

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>G01N 33/00, 24/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 00/67017</b> <b>(43) International Publication Date:</b> 9 November 2000 (09.11.00)
<b>(21) International Application Number:</b> PCT/US00/12026 <b>(22) International Filing Date:</b> 3 May 2000 (03.05.00)  <b>(30) Priority Data:</b> 09/304,799 4 May 1999 (04.05.99) US  <b>(71) Applicant:</b> THE ROCKEFELLER UNIVERSITY [US/US]; 1230 York Avenue, New York, NY 10021-6399 (US).  <b>(72) Inventors:</b> CHAIT, Brian, T.; 500 East 63rd Street, Apt. 20D, New York, NY 10021 (US). COWBURN, David; 215 East Dudley Avenue, Westfield, NJ 07090 (US). ODA, Yoshi; #102 4-8-1 Ninomiya, Tsukuba, Ibaraki 305-0051 (JP).  <b>(74) Agents:</b> SKLAR, Brandon, N. et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).		<b>(81) Designated States:</b> CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> METHOD FOR THE COMPARATIVE QUANTITATIVE ANALYSIS OF PROTEINS AND OTHER BIOLOGICAL MATERIAL BY ISOTOPIC LABELING AND MASS SPECTROSCOPY  <b>(57) Abstract</b>  The present invention is a method for accurately comparing the levels of cellular components, such as proteins, present in samples which differ in some respect from each other using mass spectroscopy and isotopic labeling. A first sample of biological matter, such as cells, is cultured in a first medium and a second sample of the same biological matter is cultured in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium. One of the samples is modulated, such as by treatment with a bacteria, a virus, a drug, hormone, a chemical or an environmental stimulus. The samples are combined and at least one protein is removed. The removed protein is subjected to mass spectroscopy to develop a mass spectrum. A ratio is computed between the peak intensities of at least one closely spaced pair of peaks to determine the relative abundance of the protein in each sample. The protein is identified by the mass spectrum or through other techniques known in the art. Modifications to the proteins, such as the phosphorylation of the protein, and the site of the modification may also be determined through the process of the present invention. The method is applicable to the components of any type of biological matter which are ionizable and may therefore be analyzed by mass spectroscopy.		

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

**METHOD FOR THE  
COMPARATIVE QUANTITATIVE ANALYSIS OF  
PROTEINS AND OTHER BIOLOGICAL MATERIAL BY  
ISOTOPIC LABELING AND MASS SPECTROSCOPY**

**Statement as to Rights to Inventions Made Under  
Federally-Sponsored Research and Development**

The U.S. Government has certain rights to the invention, by virtue of its partial support of research under National Institute of Health Grant Nos. RR00862 (BTC), GM-47021 and GM49716 (FRC).

**Field of the Invention**

The invention relates to the analysis of biological matter and, more particularly, to the comparison of isotopically labeled components of biological matter from one sample with the same, unlabeled components of biological matter from another sample, through mass spectroscopy. The method is particularly suited for quantifying differences in protein expression or modification in two cell populations or pools, one of which is subjected to environmental, genetic or chemical modulation.

**Background of the Invention**

Many biological processes in living cells are controlled by alterations in the levels or states of certain key proteins. Measuring the levels of the various proteins that affect (or are affected by) the process is therefore important for gaining an understanding of the biological process. For example, a given hormone may, through a signaling cascade, activate certain key transcription factors which in turn induce the expression of a number of proteins with distinct activities. Comparison of the levels of the proteins in the cell prior to and after induction can indicate which gene products are being up regulated and/or down regulated by the action of the hormone. As a second example, comparison of the total complement of proteins from an organism (i.e., the proteome) prior to and after infection with a virus can show which proteins are down and/or up regulated by the infection. Such an analysis can provide important information about the mechanism by which the virus

subverts its host cell, thereby aiding in the development of anti-viral drug strategies. Similarly, comparison of some or all of the proteins of the proteome before and after treatment with a drug can indicate the mechanism of action of the drug, as well as its potential effectiveness and toxicity. As another example, measurement of the state of phosphorylation of protein members of an intracellular cascade involved in turning on and off a given biological process can provide information about the control of the signaling pathway.

A facile method for accurately comparing the levels of proteins and other cellular components and biological materials as a function of time or as the result of particular treatment, such as a hormone, a drug, or a virus, as mentioned above, or an environmental stimulus, such as a temperature change, is needed. It is also necessary to assay these protein levels with high accuracy because small changes in the levels of certain key proteins may, through a complex cascade of molecular events, produce large changes in the biological system.

Two-dimensional electrophoresis has been used to compare proteins from different cell cultures or hosts subjected to differing conditions. See, for example, Anderson, N.G., et al., "Simultaneous Measurement of Hundreds of Liver Proteins: Application in Assessment of Liver Function," Toxicologic Pathology, 1996, Vol. 24, No. 1, pp. 72-76; Anderson, N.G., et al., "Twenty years of two-dimensional electrophoresis: Past, present and future," Electrophoresis, 1996, Vol. 17, pp. 443-453; Anderson, N.G., "Covalent Protein Modifications and Gene Expression Changes in Rodent Liver Following Administration of Methapyriline: A Study Using Two-Dimensional Electrophoresis," Fundamental and Applied Toxicology, 1992, Vol. 18, pp. 570-580; and Anderson, N.G., et al., "Global Approaches to Quantitative Analysis of Gene-Expression Patterns Observed by use of Two-Dimensional Gel Electrophoresis," Clin. Chem. 1984, Vol. 30, No. 12, pp. 2031-2036.

Fig. 1 is a schematic representation of the processes described in these articles. Proteins from a control cell culture are extracted, purified and separated by one- and two-dimensional electrophoresis. Proteins from another, parallel cell culture, which may include cells exposed to drugs, carcinogens or other



such treatments directly or through a host, are also extracted, purified and separated by one- and two-dimensional electrophoresis. Fig. 2 shows exemplary electrophoretic gel samples from each cell culture. Spots at different locations in each gel sample may indicate the presence of different proteins or changes in the proteins in the control versus the treated cells. Spots of different sizes may indicate a change in the quantity of the protein in the treated cells. The gels may be analyzed visually or by labeled maps, bargraphs or numerical tables. See, Anderson, "Twenty years of two-dimensional electrophoresis....", at p. 450. Computer generated arrowplots, which indicate the magnitude and polarity of changes in spots between gel samples of a control and treated cell pool, superimposed on a gel sample of a control cell pool, have also been used. Id. Instead of a control cell sample, the gel sample including the proteins from the treated cells may be compared to a master gel pattern from a library of gel patterns.

Using prior art methods, hundreds of gel samples and hundreds of thousands of protein abundance measurements may be required in a typical study. Id. It is also difficult to maintain the reproducibility of the extraction and purification procedures in each cell sample. Extraction and purification results must be normalized. Precise, accurate and reproducible quantification of the changes between cell pools is also difficult. If a gel spot includes more than one protein, the discrete proteins frequently cannot be identified. Thus, a more practical method of comparing proteins in different cell pools is needed.

Mass spectroscopy is a highly accurate analytical tool for determining molecular weights and identifying chemical structures. Proteins and peptides have been studied by matrix-assisted laser desorption mass spectroscopy and electrospray ionization mass spectroscopy. See, for example, Chait, Brian T. and Kent, Stephen B.H., "Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins", Science, Vol. 257, September 25, 1992, pp. 1885-1894, which is incorporated by reference herein. Matrix-assisted laser desorption time-of-flight mass spectrometers are described in U.S. Patent Nos. 5,045,694 and 5,453,247, to Beavis, et al., which are assigned to the assignee of the present invention and incorporated by reference herein. Electrospray ionization

mass spectrometers are described in U.S. Patent No. 5,245,186 to Chait et al., and U.S. Patent No. 4,977,320 to Chowdhury et al., for example, which are also assigned to the assignee of the present invention and incorporated by reference herein. Prior to analysis, the proteins are typically separated by one- or two-dimensional electrophoresis and then digested by an appropriate enzyme. The resulting peptides are then subjected to mass spectroscopy by any of the types of mass spectrometers identified above.

However, quantitative comparisons among proteins within a sample or between samples may be compromised by a number of parameters, such as the ionization efficiency of the mass spectrometer for a particular peptide or protein, the extraction efficiency from electrophoretic gels for a particular peptide and the digestion efficiency of an enzyme at different cleavage sites.

Isotopic labeling by stable or radioactive isotopes has been used to study many aspects of human, animal and plant metabolism. For example, isotopic labeling has been used to study metabolic turnover rates and biosynthesis of proteins and nucleic acids. Microorganisms, organs and tissue extracts, for example, may also be studied through isotopic labeling. The presence of radioactive isotopes in a sample of biological material may be detected by scintillation counters, or autoradiography, for example. However, the use of radioactive isotopes pose hazards to those conducting the experiments and require the use of protective measures, which may be cumbersome and expensive. To avoid this problem, in U.S. Patent No. 5,366,721, a long-lived radioisotope, such as carbon-14, is administered to a biological host. A reacted fraction is isolated from the host and the radioisotope concentration is measured by mass spectroscopy. See also DeLeecher, A.P. et al., "Applications of isotope dilution - mass spectrometry in clinical chemistry, pharmacokinetics, and toxicology," Mass Spectroscopy Reviews, 1992 11, 249-307; Grostic, M.F. et al., "Mass-Spectral Studies Employing Stable Isotopes in Chemistry and Biology," appearing in Mass Spectroscopy: Techniques and Applications, edited by Mike, G.W.A., Wily-Interscience (1971), pp. 217-287.

### Summary of the Invention

The present invention is a method for accurately comparing the levels of ionizable components of biological matter, wherein the biological matter differs in some respect from each other, using mass spectroscopy and isotopic labeling.

In one embodiment of the present invention, a method for comparing the relative abundance of a protein of interest in multiple samples of biological matter is disclosed, wherein one of the samples has been modulated by exposure to a treatment, such as a bacteria, virus, drug or hormone, or a stimulus, such as a chemical or environmental stimulus. A first sample of the biological matter is cultured in a first medium containing a natural abundance of isotopes and a second sample of the biological matter is cultured in a second medium containing more or less than the natural abundance of one or more isotopes. One of the samples is modulated, at least portions of the samples are combined and at least one protein is removed from the combined sample.

The removed protein, which may or may not be digested into peptides, is subjected to mass spectroscopy to develop a mass spectrum. The difference in the mass of the isotope in each cell pool results in two distinct, closely spaced peaks for each protein or peptide in the mass spectrum. One peak corresponds to a protein or peptide from a protein from the cell pool with the normal abundance of isotopes. The other peak corresponds to a protein or peptide from the cell pool enriched in one or more of the isotopes. A ratio is computed between the peak intensities of at least one pair of peaks in the mass spectrum. The relative abundance of the protein in each sample may be determined based on the computed ratio. The protein may be identified by the mass-to-charge ratios of the peaks in the mass spectrum, as well as by other means known in the art.

In addition, modifications, such as phosphorylation, glycosylation or acylation, at specific sites on individual proteins may be detected and quantified through mass spectroscopy in accordance with the present invention.

The first sample need not contain a natural abundance of isotopes, as long as at least one isotope in the second sample of biological matter has a different abundance than the abundance of the same isotope in the first sample. Preferably,

the isotope which is enriched or depleted is a non-radioactive isotope of nitrogen, oxygen, carbon and/or sulfur. Hydrogen may be used, as well. Radioactive isotopes may also be used.

The effects of two or more modulations can be simultaneously analyzed by preparing additional samples with media containing an isotope with a different abundance than the abundance of the isotope in the other samples, and modulating the additional samples.

Up to the point of the mass spectroscopy, none of the steps of the process discriminates between a protein that contains the natural abundance of isotopes from the same protein from the enriched sample. Thus, the ratios of the original amounts of proteins from the two samples are maintained, normalizing for differences between extraction and separation of the proteins in the samples.

The method is applicable to the components of any type of biological matter which can be ionized and therefore may be analyzed by mass spectroscopy. For example, the component may be a protein, a peptide, a carbohydrate, a lipid, a cofactor and post-synthetic derivatives thereof. The biological matter may be a culture of biological cells, a microbiological culture, biological tissue, an organ, an organism, a collection of organisms, a part of an organism, and a cell-free biological mimetic system, for example.

### **Brief Description of the Figures**

Fig. 1 is a flow chart of a prior art process for comparing proteins from two cell pools, one of which is diseased;

Fig. 2 shows two exemplary electrophoretic gel samples from the process of Fig. 1;

Fig. 3 is a flow chart of one embodiment of the method of the present invention;

Fig. 4 is an exemplary mass spectrum resulting from a hypothetical experiment conducted in accordance with the method of Fig. 3;

Fig. 5 is a portion of a mass spectrum of Abl-SH2;

Fig. 6 is an enlarged view of the region 6 of the mass spectrum of Fig. 5;

Fig. 7 is a graph of measured ratios between the peak intensities corresponding to peptides of Abl-SH2 from an unlabeled (nitrogen-14) sample and the peak intensities corresponding to a labeled (nitrogen-15) sample, versus the expected ratios, based on seven spectral measurements of each of the samples of Table II.

Fig. 8 is a portion of a mass spectrum of the peptide of the protein elongation factor 1- $\alpha$  from a combined pool of labeled and unlabeled *Saccharomyces cerevisiae* which differed only in their ability to express the cyclin *CLN2*;

Fig. 9a is an enlarged view of region 9a in the mass spectrum of Fig. 8, showing a pair of peaks from a single peptide;

Fig. 9b shows the theoretical isotope distributions of nitrogen-14 and nitrogen-15;

Fig. 10 is a portion of a mass spectrum of the peptides of the protein triosephosphate isomerase from the combined cell pool described above with respect to Fig. 8;

Fig. 11 is a schematic representation of a mass spectrum, indicating a plurality of pairs of peaks, including a pair of peaks corresponding to a unphosphorylated peptide (X) and a pair of peaks corresponding to the same peptide, which has been phosphorylated ( $X_p$ );

Fig. 12 is a bar graph showing the changes in the ratios of the unphosphorylated peptide X and the phosphorylated peptide  $X_p$ ; and

Fig. 13 shows the experimentally observed change of the intensity ratio of the unphosphorylated peptide as well as the change observed for the singly, doubly and triply phosphorylated Ste20 peptide obtained from *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup> cells.

### **Detailed Description of the Invention**

Fig. 3 is a flow chart of one embodiment of the process of the present invention for comparing the quantities of proteins in different cell samples. Two cell pools are prepared. (Step 10). One of the cell pools, here Cell Pool 1, contains a cell culture grown or maintained in a medium containing a natural abundance of

isotopes. For example, the medium contains 99.6% nitrogen-14 and 0.4% nitrogen-15, which are the naturally occurring abundances of those isotopes of nitrogen. The isotopes of oxygen, carbon, sulfur and other elements are also present in their naturally occurring abundances.

The other cell pool, here Cell Pool 2, contains a cell culture grown or maintained in a medium in which one or more isotopes of nitrogen, carbon, oxygen or sulfur, for example, is not present in a natural abundance. For example, the second medium may be isotopically enriched in one or more of the following isotopes: nitrogen-15, carbon -13, oxygen -17, oxygen -18, and sulfur -34.

Enrichment of hydrogen-2 (deuterium) can also be used, however, the enzymatic effects of hydrogen substitution is greater than that of  $^{15}\text{N}$  and can therefore change the biological process.

High enrichment is preferred. The particular level of enrichment may depend on the isotope chosen. For example, a medium enriched in nitrogen -15 ( $^{15}\text{N}$ ) to between 90% - 100% is preferred, with 100%  $^{15}\text{N}$  enrichment most preferred. Because hydrogen -2 can be toxic, less than 100% hydrogen -2 is preferred. Isotopic depletion may also be used wherein less than the naturally occurring abundance of an isotope is provided in the medium. Non-radioactive isotopes are preferred. Radioactive isotopes may also be used, but are not preferred because of the difficulties the use of such isotopes present, as discussed above.

The first cell pool need not contain a natural abundance of isotopes, as long as at least one isotope in the second cell pool has a different abundance than the abundance of the same isotope in the first cell pool.

It is preferred that the media the cell pools are grown in be identical, except for the presence of the different isotope or isotopes. Bio-Express from Cambridge Isotope Laboratories, Inc., Andover, Massachusetts, is one suitable  $^{15}\text{N}$  enriched media. Alternatively, the cells may be grown in the same medium and the labeling isotope may be added directly to one culture of cells.

Either of the cell pools, here Cell Pool 2, is modulated by a bacteria, a virus, a chemical, a drug, a hormone, or an environmental change, such as a

temperature change, for example. Other treatments or stimulus may be provided, as well. The other cell pool, here Cell Pool 1, acts as a control.

All or portions of the cell pools are then combined. (Step 20). The proteins are extracted from the combined cell pool in a manner known in the art. (Step 30). For example, the cell membranes may be digested or disrupted by standard methods, such as detergents or homogenization in an isotonic sucrose solution. The proteins are then extracted from the combined cell pools by ultra-centrifugation, or other known techniques. For example, antibodies may be used to immunoprecipitate certain proteins or complexes of proteins, as well. The particular method used may be dependent on the particular proteins of interest, as is known in the art.

The mixture of proteins is then separated into the individual proteins or small groups of proteins, also by known techniques, such as one- and/or two-dimensional electrophoresis, ultra-centrifugation, chromatography or affinity binding. (Step 40). Two-dimensional sodium dodecylsulfate-polyacrylamide gel electrophoresis ("SDS-PAGE"), may be used, for example. If an individual protein is extracted from the combined cell pool, such as by use of an antibody, the separation step 40 may not be necessary.

The separated proteins are then preferably digested into peptides. (Step 50). Preferably, the proteins are digested by a proteolytic enzyme. Trypsin is preferred because it cleaves precisely at the sites of lysine and arginine, yielding doubly-charged peptides which typically have a length of from about 5 to 50 amino acids and a molecular weight of between about 700-5,000. Such peptides are particularly appropriate for analysis by mass spectroscopy, especially by electrospray ionization mass spectroscopy. Other site specific proteolytic enzymes which may be used include Ly-C, Asp-N and Glu-C, for example. Pepsin, subtilisin and proteinase 1c are low specificity enzymes which may also be used. Chemical reagents may also be used to digest the proteins. For example, cyanogen bromide may be used to cut a protein into peptides at the site of methionine. BNPS-skatole may be used to cleave at the site of tryptophan. Acid hydrolysis may also be used.

o The proteins or digested proteins are then subjected to mass spectroscopy. (Step 60). Any mass spectrometer may be used to analyze the peptides or proteins. For example, the mass spectrometer may be a Matrix-Assisted Laser Desorption/Ionization ("MALDI") Time-of-Flight ("TOF") Mass Spectrometer, available from PerSeptive Biosystems, Framingham, Massachusetts; 5 an Electrospray Ionization ("ESI") ion trap mass spectrometer, available from Finnigan MAT, San Jose, California; or an ESI quadrupole mass spectrometer, available from Finnigan MAT or the Perkin-Elmer Corporation, Foster City, California.

10 A simple mixture of from 1 to about 5 digested proteins can be analyzed by single-stage mass spectroscopy with any of the mass spectrometers discussed above. Mixtures of greater than six digested proteins are preferably analyzed by a two-stage tandem mass spectroscopy procedure involving collision 15 produced dissociation ("CID"), as is known in the art.

While preferred, the digestion step 50 is not required. One or several whole proteins can also be subjected to mass spectroscopy, avoiding the need for digesting the proteins into peptides, as is known in the art. Single-stage mass 20 spectroscopy may be used to analyze mixtures of large numbers of whole proteins simultaneously.

The protein or proteins subjected to the mass spectroscopy are also preferably identified. (Step 70). The identification step 70 can take place at any time after separation or extraction of a single protein. Protein identification 25 software which uses algorithms to compare the mass spectrum with a database of proteins are available. One such algorithm, ProFound, uses a Bayesian algorithm to search protein or DNA databases to identify the optimum match between the experimental data and the protein in the database. W. Zhang, B.T. Chait, 30 "Proceedings of the 43<sup>rd</sup> ASMS Conference on Mass Spectroscopy and Allied Topics," Atlanta, Georgia (1995) p. 643. ProFound may be accessed on the World-Wide Web at <<http://prowl.rockefeller.edu>> and <<http://www.proteometrics.com>>. ProFound accesses the non-redundant database (NR). Alternative algorithms for 35 protein identification include: Mass Search (<http://cbrg.inf.ethz.ch/subsection>



3\_3.html); MOWSE (<http://www.seqnet.dl.ac.uk//mows.html>); MSFIT  
(<http://prospector.ucsf.edu/ucsfhtml/msfit.htm>); Peptide Mass Search  
([http://www.mdc-berlin.de/~emu/peptide\\_mass.html](http://www.mdc-berlin.de/~emu/peptide_mass.html)); and Peptide Search  
([http://www.mann.embl\\_heidelberg.de/services/peptide\\_search/fr\\_peptide  
searchform.html](http://www.mann.embl_heidelberg.de/services/peptide_search/fr_peptide_searchform.html)). See also, James, Peter, "Protein identification in the post-genome  
era: the rapid rise of proteomics", Q. Rev. Biophysics, Vol. 30, No. 4, pp. 279-331  
(1997), which is incorporated by reference, herein.

The protein can also be identified by electrophoresis, antibodies  
Edman sequencing or bioassay, or by other methods conventionally used in the art,  
after separation of the proteins in step 40.

The ratios of the peak intensities of each pair of peaks are then  
computed. (Step 80). The ratios give a measure of the relative amount of that  
peptide in each cell pool, as discussed further, below. The peak intensities are  
calculated in a conventional manner.

Fig. 4 is an exemplary mass spectrum of four peptides from a single  
protein resulting from a hypothetical experiment conducted in accordance with the  
method of Fig. 3.

Because of the difference between the masses of the peptides or  
proteins and resulting peptides from cells grown in the non-labeled medium and  
those grown in the labeled, isotopically enriched medium, the results of the mass  
spectroscopy will generally be a plurality of pairs of closely spaced peaks, each peak  
being at a different  $m/z$  ratio. Since the enriched isotope is typically heavier than  
the most abundant naturally occurring isotope ( $^{15}\text{N}$  versus  $^{14}\text{N}$ , for example), the  
peak at the higher  $m/z$  ratio is generally indicative of the relative abundance of the  
peptide from a labeled protein grown in the medium enriched in one or more  
isotopes. The peak at the lower  $m/z$  ratio is generally indicative of the relative  
abundance of the peptide from an unlabeled protein in the medium containing a  
normal abundance of isotopes.

Table I, below, lists four ratios of the peak intensities of isotopically  
labeled peptides of Cell Pool 2 and non-isotopically labeled peptides of the control  
Cell Pool 1 ( $X_i^2/X_i^1$ ) based on the exemplary mass spectrum of Fig. 4.

TABLE I

	Peptide No.	Intensity Ratio
	1	$0.70 \pm .02$
	2	$0.69 \pm .02$
	3	$0.30 \pm .02$
5	4	$0.71 \pm .02$

Since the number of cells in one cell pool may differ from the number of cells in the other cell pool, for any given pair of peaks, the intensity of a peak corresponding to a peptide from one cell pool may differ from the intensity of the peak corresponding to the same peptide from the other cell pool. The ratios between most of the pairs of peaks (which are indicative of peptides derived from proteins unaffected by a given treatment or stimulus), will generally be the same. A deviation from the regularly-observed ratio indicates a difference in the relative quantity of a peptide, and hence a protein, between the two cell pools which may be caused by the modulation to which one of the cell pools has been subjected. The difference can be quantified in accordance with the present invention.

Because the cell pools are combined, other sources of differences in the intensities of the peaks, such as variations in the extraction efficiency of a particular protein from the cell pool, the subsequent extraction efficiency of a protein from the gel, the digestion efficiency of the enzyme used (if any), the ionization efficiency of the mass spectrometer for a particular peptide, and other such factors, affect both cell pools equally. These factors should not, therefore, affect the observed ratios. Analyzing the ratio of the pair of peaks compensates for differences in mass intensities resulting from differences in the ionization efficiency of the mass spectrometer for a particular peptide. Isotopically labeling one of the two cell pools and observing the ratio between the peaks of the isotopically labeled and non-isotopically labeled peptides also compensates for differential effects between the cell pools themselves, such as the presence of a different number of cells in each, providing an internal normalization between the cell samples.

To ensure that the change in the ratios is not caused by the isotopic enrichment itself, the process of Fig. 3 is preferably repeated with the other cell pool

being isotopically enriched, *i.e.*, if in the first run the treated cell pool is isotopically enriched, as in Fig. 3, then in the second run, the control cell pool would be isotopically enriched.

The ratios of the intensities of the peaks in the hypothetical spectrum shown in Fig. 4 is about 0.70, except for peptide number 3, whose ratio is about 0.30. This indicates that the relative amount of the protein corresponding to peptide 3 in the treated and control cell pools is different than the relative amounts of the protein or proteins corresponding to peptides 1, 2 and 4. This suggests that the modulation affects the post-translational expression of the protein from which peptide 3 is derived. In an actual example, there would be other pairs of peaks having ratios of about 0.30, corresponding to other peptides from the protein from which peptide number 3 is derived. Analysis of the plurality of peaks having a ratio of about 0.30 would enable an identification of the protein through the protein identification algorithms discussed above.

The percentage difference between the regularly observed ratio of the peak intensities of the peptides from both cell pools and the observed ratio for the modified peptides, *e.g.*, peptide 3, in Fig. 4, is the percentage change in the expression of the protein in the treated cell pool, *e.g.*, Cell Pool 2, which could be caused by the disease or another type of treatment or stimulus. In this example:

$$\frac{0.30 - 0.70}{0.70} \times 100 = -57\%$$

This suggests that the disease decreased the expression of the protein in the modulated Cell Pool 2 by 57%, compared with the expression of the same protein in the control Cell Pool 1. This quantitative change caused by the disease may give insight into the effect of the disease or other such modulation on the cell.

The method of the invention can also be extended to comparing the effects of two or more modulations by preparing three or more cell pools wherein at least one isotope in each cell pool is different. For example, if the proteins in three cell pools are to be compared, preferably one cell pool, which in this example is the non-modulated cell pool, would have a normal abundance of isotopes, the second cell pool would be enriched in  $^{15}\text{N}$  and the third cell pool would be enriched in a different isotope, such as carbon-13, as well as  $^{15}\text{N}$ . The resulting mass spectrum

- 14 -

would then comprise groups of three closely spaced peaks, each corresponding to a protein or peptide from a different cell pool. The ratios between the peak intensities of the proteins or peptides from each of the modulated cell pools and the non-modulated cell pool would be indicative of the relative amount of that peptide in each cell pool.

### **Example 1**

To demonstrate the internal normalization provided by mixing isotopically enriched and non-enriched proteins, the following experiment was conducted. No modulation was applied to either cell pool in this example.

Human Abelson protein tyrosine kinase Src homology domain-2 (Abl-SH2) was derived from E. coli DH $\alpha$ 5 strain cells grown in minimal medium M9, with 1 gram per liter of  $^{15}\text{NH}_4\text{Cl}$  (ammonium chloride including  $^{15}\text{N}$ ) as the sole source of  $^{15}\text{N}$ . The preparation of labeled and unlabeled Abl-SH2 from E. coli DH $\alpha$ 5 is described in Gosser, Y.Q., et al., Structure 3, 1075 (1995), which is incorporated by reference, herein. The following samples were prepared:

**TABLE II**

Sample Number	1	2	3	4	5	6	7
Natural Protein (A) picomole/microliters	0.1	0.25	0.5	1.0	1.5	2.0	2.5
$^{15}\text{N}$ Labeled Protein (B) picomole/microliters	1.5	1.5	1.5	1.5	1.5	1.5	1.5

The natural samples were prepared by dilution into 50mM ammonium carbonate ( $\text{NH}_4\text{HCO}_3$ ). 10 microliters of the natural and labeled proteins were combined and subjected to SDS-PAGE with 4-20% acrylamide concentration tris-glycine gel, available from Novex, San Diego, California, in the form of pre-cast mini-gels. A voltage of 130 volts was applied for two hours. After two hours, the gel was stained with Copper Stain from Bio-Rad Laboratories, Hercules, California.

Each protein spot or band was cut out of the gel and destained twice, for 10 minutes each. The gel pieces were then washed twice, for 15 minutes each. The gel pieces were then crushed.

0.2 micrograms of trypsin, from Boehringer Mannheim, and 50 millimolar  $\text{NH}_4\text{HCO}_3$  (ammonium carbonate) having a pH of 8 were added to the crushed gel pieces. The mixture was allowed to stand for 2 hours at  $37^\circ\text{C}$ . The peptides were then removed from the gel pieces by acetonitrile ( $\text{CH}_3\text{CN}$ ) at a concentration of 75%.

The solution was dried in a SpeedVac, available from Savant, Holbrook, NY. The lyophilized material was then dissolved in 7 microliters of about 50%  $\text{CH}_3\text{CN}$ , about 50%  $\text{H}_2\text{O}$  and about 0.1% trifluoroacetic acid ("TFA"). 10% of the sample solution, 0.7 microliters, was mixed with 0.7 microliters of 2-5 dihydroxybenzoic acid (gentisic acid) referred to as MALDI-MATRIX DHB, at a concentration of 60 grams per liter. MALDI-MATRIX DHB is available from Aldrich Chemical Company, Inc., Milwaukee, Wisconsin. The solution was then analyzed with a Model STR Matrix-Assisted Laser Desorption/Ionization time-of-flight mass spectrometer ("MALDI-TOF-MS"), available from PerSeptive Biosystems.

A portion of the resulting mass spectrum is shown in Fig. 5. Several pairs of peaks 110, 120, 130, 140 and 150 are shown. In each pair, the peak of the higher mass-to-charge ratio (" $m/z$ ") is indicative of a peptide of Abl-SH2 from the cell pool enriched with  $^{15}\text{N}$ , while the peak at the lower  $m/z$  ratio is indicative of a peptide of the Abl-SH2 from the non-enriched cell pool, which is predominantly  $^{14}\text{N}$ . For example, in peak pair 120, peak A is indicative of a peptide from the unlabeled cell pool while peak A' is indicative of the same peptide from the labeled cell pool. Fig. 6 is an enlarged view of the region 6 of the mass spectrum of Fig. 5, including the peaks 120 and 130 in the mass-to-charge ratio ( $m/z$ ) range of about 940-990. The corresponding peaks from Fig. 5 are similarly labeled in Fig. 6. The closely bunched multiple peaks in Fig. 6 are due to the naturally occurring distribution of the various isotopes of nitrogen, carbon, oxygen, sulfur and hydrogen in the peptides.

- 16 -

The amino acid sequence of Abl-SH2 appears below:

	1	10	20	30	40	50
1	GSGNSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQRSISLR					
51	YEGRVYHYRINTASDGKLYVSSSRFNTLAELVHHHSTVADGLITTLHYP					
101	APKRGIHRL [SEQ. ID NO: 1]					

The pair of peaks 110 m/z ~ 737 in Fig. 5 corresponds to the peptide sequence starting with the amino acid number 55 and ending with 59; the pair of peaks 120 at m/z ~ 940 corresponds to the peptide sequence spanning amino acid numbers 68-75; the pair of peaks 130 at m/z ~ 976 corresponds to the peptide sequence spanning amino acid numbers 37-45; the pair of peaks 140 at m/z ~ 1225 corresponds to the peptide sequence spanning amino acid numbers 9-18; and the pair of peaks 150 at m/z ~ 1910 corresponds to the peptide sequence spanning amino acid numbers 19-36.

The average peak intensity ratio and relative standard deviation ("RSD") for each of 5 peptides from the digested Abl-SH2 in 7 individual mass spectral measurements are shown below:

TABLE III

## Peak Intensity Ratio

Theoretical Loading Amount of  $^{14}\text{N}$  protein -1 picomole ("pm")  
( $^{15}\text{N}$  protein amount - 1.5pm)

Peptide	m/z (M+H)	Peptide Sequence	Average Ratio (n=7)	R.S.D.
1	737	55-59	0.61	6.9%
2	940	68-75	0.60	5.6%
3	976	37-45	0.67	13.6%
4	1225	9-18	0.62	3.1%
5	1910	19-36	0.61	6.0%
5 peptides total			0.62	4.8%

o The average of the RSD for the seven spectral measurements of the five peptides was 4.8%.

Fig. 7 is a graph of the measured ratios between the peak intensity corresponding to unlabeled  $^{14}\text{N}$  and the peak intensity corresponding to labeled  $^{15}\text{N}$ , based on seven spectral measurements of each of the samples of Table II, above. Since no modification was applied to either cell pool, the only difference between the samples in this example was the concentration of the natural protein in each cell pool. The value of the intensity ratios of each of the pairs of peaks for each peptide should therefore be equal to the ratio of the concentrations of the protein Abl-SH2 in each cell pool. The intensity ratio was found to be linear ( $R=0.997$ ) over an abundance ratio of more than 10:1.

### Example 2

In another example, the levels of high abundance proteins derived from two pools of *Saccharomyces* ("S.") *cerevisiae* cells that differed only in their ability to express the G1 cyclin *CLN2* was analyzed. *CLN2* is important in regulating the G1-S transition in budding yeast, but the effect of its expression on the levels of specific proteins and their modification is largely unknown. See F.R. Cross, *Curr. Opin. Cell Biol.* 6, 790 (1995). The cells in one cell pool were *cln1 cln2*, mutant yeast harboring a GAL1::CLN2 overexpression cassette. The cells in the second cell pool were yeast without the overexpression cassette. Both populations were proliferating but only one population was expressing CLN2. The expressing population is designated "*CLN2<sup>+</sup>*". The non-expressing population is designated "*cln2<sup>-</sup>*".

A first combined sample contained 1 ml of unlabeled ( $^{14}\text{N}$ ) extract of *cln2<sup>-</sup>* plus 1 ml of  $^{15}\text{N}$ -labeled extract of *CLN2<sup>+</sup>*. A second combined sample contained 1 ml of unlabeled ( $^{14}\text{N}$ ) extract of *CLN2<sup>+</sup>* plus 1 ml of  $^{15}\text{N}$ -labeled extract of *cln2<sup>-</sup>*. These two different samples were prepared to control for systemic errors in the measurements.

The enriched media was Bio-Express-1000 (discussed above), specified by the manufacturer as enriched to greater than 96%  $^{15}\text{N}$ . The enriched

- 18 -

- o and non-enriched media were both glucose-free and were supplemented with unlabeled tryptophan. Galactose was added to each media to 0.3%. The cells were grown overnight to mid-log phase ( $OD \leq 1.0$ ) at 30° C with shaking.

The combined samples were separated by a combination of reversed phase high performance liquid chromatography ("HPLC") and SDS-PAGE. The HPLC column was 10 mm x 100 mm, and contained C4 silica gel (Brownlee Prep-10 Butyl, 20  $\mu$ m, 300 Å, Perkin Elmer, Norwalk, CT). The HPLC mobile phase A was composed of 67% ACS grade formic acid and 33% water. The mobile phase B was composed of 67% formic acid and 33% acetonitrile. The flow-rate was 1 ml/min. The gradient curve was 0 - 5 min B conc. 0%; 5 min - 10 min B conc. from 0 - 30%; 10 min - 110 min B conc. from 30 - 100%; 110 min - 120 min B conc. was 100%.

A total of 57 fractions of 2 ml each were collected. The HPLC injection volume was 14 ml, composed of 2 ml of whole yeast extract and 12 ml of mobile phase A.

After HPLC, the proteins were precipitated from each 2 ml fraction by adding 10 ml of water, 1 ml of 0.4% deoxycholate and 1 ml trichloroacetic acid (1g/ml). The solution was maintained at 4°C for 1 hr and then centrifuged at 2,200 x g for 30 min at 4°C. The supernatant was then discarded. One milliliter of 80% acetone was added and the sample was transferred to a new Eppendorf tube. This sample was maintained at -20°C for 1 hr before centrifuging at 14,000 x g for 5 min at room temperature. The supernatant was discarded and aqueous NaOH was added for neutralization.

SDS-PAGE sample buffer was added and each fraction was run on a separate lane of an 8-16% gradient tris-glycine gel at constant voltage of 110 V. The gels were stained with colloidal Coomassie Brilliant Blue. The gradient tris-glycine gel and Coomassie Brilliant Blue are available from Novex, for example.

The protein bands were excised from the gels with a scalpel and placed into Eppendorf tubes. The gel pieces were destained and washed until clear with methanol/water/acetic acid 5:4:1 (v/v/v) with 3X exchange of solution,



followed by a 30 min vortex in high purity water. The water was discarded and the gel sliced into 2-3 mm pieces, soaked in 0.5 ml of acetonitrile, and vortexed.

After the gel pieces had shrunk and turned opaque, the acetonitrile was discarded. A trypsin solution of 0.2 micrograms ("µg") in 50mM  $\text{NH}_3\text{HCO}_3$  was added in sufficient volume, 20-50 microliters ("µl"), to re-hydrate the gel pieces. After incubation for 2 hours at 37°C, 30 µl of acetonitrile was added to the gel pieces and vortexed for a few minutes. A further 30 µl of acetonitrile was added and the process repeated until the gel turned opaque white. The supernatant was transferred to a new tube, the gel was rehydrated with water (30 µl) and the extraction steps repeated. The supernatant was dried using a SpeedVac and each dried sample was re-dissolved in 5 µl of acetonitrile/0.1% aqueous trifluoroacetic acid ("TFA") 1:2 (v/v). 0.5 µl of this sample solution was loaded onto the sample plate together with 0.5 µl of matrix solution (DHB: 2,5-dihydroxybenzoic acid). Mass spectrum measurements were obtained using a MALDI-TOF-MS Model STR, from PerSeptive Biosystems, Framingham, MA, operated in reflector mode.

The abundances of a selection of individual proteins from the two samples were then compared. Figs. 8 and 10 are examples of MALDI-TOF-MS mass spectra of tryptic peptides from the combined cell pools of unlabeled  $^{14}\text{N}$  *cln2*<sup>-</sup> and labeled *CLN2*<sup>+</sup>. The peptides in the mass spectrum of Fig. 8 originate from one single protein while the peptides in the mass spectrum of Fig. 10 originate from another single protein. These proteins were isolated from several hundred separated by the combination of reversed-phase HPLC and SDS-PAGE, as described above. Fig. 8 shows pairs of peptide peaks 160, 162, 164 and 166. Fig. 10 shows pairs of peptide peaks 168, 170, 172, 174 and 176. Peaks 160a, 162a, 164a and 166a in Fig. 8 and peaks 168a, 170a, 172a, 174a and 176a in Fig. 10 arise from the unlabeled ( $^{14}\text{N}$ ) *cln2*<sup>-</sup> protein. Peaks 160b, 162b, 164b, 166b in Fig. 8 and peaks 168b, 170b, 172b, 174b and 176b in Fig. 10 arise from  $^{15}\text{N}$  labeled *CLN2*<sup>+</sup> protein.

Fig. 9a is an enlarged view of one pair of peaks 160 in region 9a of the mass spectrum of Fig. 8. The cluster of peaks 160a corresponds to isotopically resolved components of the unlabeled peptide while the cluster 160b corresponds to the isotopic components of the  $^{15}\text{N}$  labeled peptide. Fig. 9b shows the theoretical

isotope distribution of  $^{14}\text{N}$  (160a') and  $^{15}\text{N}$  (160b'). Tests of the goodness of fit of the theoretical isotope distribution (FIG. 9b) to the experimental distribution (FIG. 9a) revealed that the level of incorporated  $^{15}\text{N}$  was  $93 \pm 1\%$ . The multiple pairs of peaks within a spectrum provide multiple measurements of the relative abundance for each identified protein. For each protein subjected to mass spectroscopy, the ratio of the abundance of the protein in the two cell pools was obtained from the intensity ratios of the pairs of peaks in the corresponding spectrum of tryptic peptides by comparing the sum of the intensities of the isotopically resolved components of the unlabeled peptide with the corresponding sum from the  $^{15}\text{N}$  labeled peptide.

The sets of masses of the lower mass components of each pair (i.e., sets of tryptic peptides from the unlabeled proteins) were used to identify the proteins from the *S. cerevisiae* database with the protein identification algorithm ProFound, discussed above. The spectrum of Fig. 8 was found to originate from the protein "elongation factor 1- $\alpha$ " ("EF1- $\alpha$ ") while the spectrum of Fig. 10 was found to originate from the protein "triosephosphate isomerase" ("TIM"). The ProFound probability value ("P") for EF1- $\alpha$  was 1.00, readily discriminating against the second ranked choice of clb5 ( $p=9.0^{-20}$ ). The ProFound probability score for TIM was 1.00, while the second ranked choice was ORF YDL100c ( $P=1.0^{-16}$ ).

The relative abundance of a selection of proteins in the *cln2*<sup>+</sup> versus CLN2<sup>+</sup> cell pools determined from the ratio of labeled ( $^{15}\text{N}$ ) versus unlabeled ( $^{14}\text{N}$ ) peptide mass spectrometric peak intensities is shown in Table IV, below.

Measurement of 42 high abundance yeast proteins revealed that these ratios fall into two categories. The first category, which includes the majority of the proteins studied, yielded intensity ratios that are the same to within the relative experimental error ( $\pm 10\%$ ) of the measurement. The average of this category of intensity ratios were normalized to 1.00 on the assumption that they arise from proteins whose relative abundances do not change in the two cell pools. Elongation factor 1- $\alpha$  falls into this first category. The second category arises from proteins whose relative abundance differs in a statistically significant manner (Standard Deviation  $> 3$ ) from the first category. Triosephosphate isomerase falls into this second category because its unlabeled-to-labeled peak ratio was determined to be 0.58. Only two

- 21 -

- other proteins of the 42 sampled, a putative peroxisomal membrane protein, ORF YLR109w (4) (ratio 0.67) and S-adenosylmethionine synthetase 2 (Sam2) (ratio 0.70), were observed to fall in this second category.

TABLE III

5	Molecular Mass (kDa)		Gene	Protein Name	$\frac{cln2^- (^{14}\text{N})}{CLN2^+ (^{15}\text{N})}$	$\frac{cln2^- (^{15}\text{N})}{CLN2^+ (^{14}\text{N})}$
	Meas	Calc				
10	23	21.6	tsa1	Thiol-specific antioxidant protein	0.89	0.80
	27	26.7	tpi1	Triosephosphate isomerase	0.58	0.59
	29	27.5	gmp1	Phosphoglycerate mutase 1	1.08	1.14
15	34	34.8	bel1	Guanine nucleotide binding protein	1.10	1.08
	37	35.6	tdh3	Glyceraldehyde 3-phosphate dehydrogenase 3	1.12	1.04
	45	44.7	pgk1	Phosphoglycerate kinase	0.98	1.09
20	49	46.7	eno2	2-phosphoglycerate dehydratase	0.98	0.94
	51	49.9	tef1	Elongation factor 1- $\alpha$	1.00	0.84
	60	54.5	cdc19	Pyruvate kinase 1	1.12	0.97
25	90	93.3	ef1	Elongation factor 2	0.91	1.00
	110	111	kgd1	$\alpha$ -ketoglutarate dehydrogenase	0.96	1.11
	120	116	yef3	Elongation factor 3	0.86	1.00

30 In Table III, Column 1 "Meas" is the molecular mass of the peptide as determined by SDS-PAGE, Column 2 "Calc" is the calculated molecular mass and Column 3 "Gene" is the Yeast Protein Database (4) gene name.

35 The normalized intensity ratios from the  $cln2^- (^{15}\text{N})/CLN2^+ (^{14}\text{N})$  (Column 6) cell pools were found to be in agreement with those from

*cln2*<sup>-</sup> (<sup>14</sup>N)/*CLN2*<sup>+</sup> (<sup>15</sup>N) (Column 5) pools to within the statistical uncertainty of the measurement, as shown in Table III, above.

Additional examples of proteins by Yeast Protein Database gene name that were identified from the two cell pools together with their abundance ratios *cln2*<sup>-</sup> (<sup>14</sup>N)/*CLN2*<sup>+</sup> (<sup>15</sup>N) in parentheses are: *ilv5* (0.93); *grs1* (0.97); *acs2* (0.99); *por1* (1.06); *pfk2*; *cdc48*; *gdh1*; *pfk2* (0.95); *cdc48* (0.90); *gdh1* (1.06); *pet9* (1.09); *pdc1* (1.00); *YLR109w* (0.67); *fba1* (1.18); *asn2* (0.97); *cys4* (1.05); *idh1* (1.10); *rpl5* (0.99); *efb1* (0.96); *YKL056c* (1.14); *tif51a* (0.94); *rps5* (1.16); *act1* (1.12); *hxx2* (1.11); *pgi1* (1.19); *ssa1* (0.95); *fas1* (0.99); *hsc82* (0.84); *hom6* (0.96); *rpp0* (1.03); *rnr2* (0.91); *bmh1* (0.91); *sam2* (0.70).

Through the process of the present invention, it was demonstrated that the abundance of the triosephosphate isomerase, ORF *YLR109w* and S-adenosylmethionine synthetase 2 (*Sam2*) was less in the *cln2*<sup>-</sup> sample than in the *CLN2*<sup>+</sup> sample, while the abundance of the elongation factor 1- $\alpha$  was the same in both samples. The biological implications of the change in the abundances of these proteins remain to be elucidated.

Based on a statistical analysis of the results, it is believed that changes in the abundance of proteins outside of two standard deviations (>20%) can be readily discerned.

The method of the present invention can also be applied to the detection of changes in the levels of protein components of incompletely separated mixtures provided that individual peptides in the MALDI-MS peptide map can be unambiguously assigned to specific proteins. Thus, for example, two 75 kDa proteins, glycyl- tRNA synthetase and acetyl coenzyme A synthetase 2, were identified by the algorithm ProFound, discussed above, which also automatically identifies proteins present as binary mixtures, and their abundance changes determined from a single gel band. These same identifications were made by subjecting a portion of the same sample from the 75 kDa band to HPLC separation followed by online electrospray ionization mass spectrometry ("ESI-MS") and ESI-tandem mass spectroscopy analysis using an ion trap mass spectrometer. D. Ashcroft et al., Electrophoresis 19, 968 (1998). The tandem mass spectroscopy

o (“MS/MS”) experiment identified proteins from the fragmentation patterns of individual peptides using the search algorithm PepFrag, which can also be accessed on the World-Wide Web at the same cite as ProFound, identified above. The corresponding MS experiment was used to obtain the intensity ratio for quantification. This combined liquid chromatography mass spectroscopy (“LC-MS”) and liquid chromatography tandem mass spectroscopy (“LC-MS/MS”) approach should also allow for the quantitative analysis of even more complex mixtures of proteins.

### 10 Site Specific Modifications

The method of the present invention also enables the identification of the site (peptide) of a modification or modifications of a protein and quantification of the difference in the degree of the modification of the protein in the control and treated cell pools, which may give insight into the mechanism of a disease or other cellular process. A hypothetical example is discussed below.

15 A control cell pool and an isotopically labeled, diseased cell pool are prepared and combined and the proteins extracted and separated, as described above. A single protein of interest is removed from the gel, digested and subjected to mass spectroscopy, also as described above. Fig. 4 and Table I demonstrate exemplary results for this example, as well.

Peptides from the two cell pools that either remain unmodified or do not undergo a change in the level of modification yield pairs of peaks with a fixed ratio of intensities -- a ratio that can be used to normalize the amounts of the protein from the two cell pools. By contrast, peptides that undergo a change in their level of modification yield pairs of peaks with intensity ratios that reflect these changes.

25 In Fig. 4, the ratios between the intensities of each pair of peaks corresponding to the same peptide from each cell pool is substantially the same, except for peptide 3. This deviation from the norm could be indicative of a change in the peptide in the corresponding protein in the diseased cell pool, which could give insight into the mechanism of the disease. For example, such a deviation could be indicative of greater or less phosphorylation, glycosylation, acylation, etc. of that peptide in the diseased cell pool than in the normal cell pool.

The m/z values of the pair of peaks in the mass spectrum is indicative of the type of change in the peptide, *i.e.*, whether a phosphate, carbohydrate or other such group is bonded to the peptide or not.

Two cases are considered. In the first case, the m/z ratio indicates that no modification of the peptide has taken place, for example, the peptide has not been phosphorylated. In the second case, the m/z ratio indicates that the peptide has been modified, such as by being phosphorylated. The percent difference between the regularly observed ratio and the observed ratio for peptide 3 in the second case is the change in the percentage of those peptides in the diseased cell pool which have been modified compared with the percentage of those peptides which have been modified in the control cell pool, here:

$$\frac{0.30 - 0.70}{0.70} \times 100 = -57\%$$

In this example, fifty seven percent of the peptide 3 which was present in the control cell pool has been modified in the diseased cell pool. Hence, 43% of the peptide 3 which was present in the control cell pool has not been modified in the diseased cell pool.

Conclusions can also be drawn as to the level of the state of modification, as shown in Example 3, below.

### **Example 3**

In this example, the degree of phosphorylation of PAK-related Sterol 20 protein kinase ("STE 20") in normal and mutant yeast cells during a signaling cascade were compared.

Cln2-Cdc28 cyclin-dependent kinase inhibits the mating factor signal transduction pathway by interfering with the function of Ste20, which correlates with Cln2-dependent *in vivo* phosphorylation of Ste20. See, for example, Oehlen, L.J.W.M., Cross, F.R., Genes Dev. 8, 1058 (1994); Oehlen, L.J.W.M., Cross, F.R., J. Biol. Chem. 273, 25089 (1998); Wu C, Leeuw T, Leberer E, Thomas DY, Whiteway M., J. Biol. Chem. 273, 28107 (1998). The differences in phosphorylation of Ste20<sub>trunc</sub> in *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup> cell pools were monitored in

- 25 -

accordance with the method of the present invention to identify Cln2<sup>+</sup>-dependent *in vivo* phosphorylation sites in Ste20.

Plasmid pYGEX-STE20 (B3553) expresses GST-Ste20 from the *GAL1* promoter, as described in R.L. Roberts, et al., *Cell* 89, 1055-65 (1997). GST-Ste20 phosphorylation site mutants were amplified via PCR from pVTU-STE 20-based constructs using an internal oligonucleotide upstream of the BamHI site and a 3' oligo which hybridized outside of the multiple cloning site of the plasmid. The 3' oligo added an SpeI site, and PCR products were cleaved with BamHI and SpeI and transferred to B3553 cut with BamHI and XbaI to create wild type and mutant STE20 alleles.

To prepare the GST-Ste20<sub>trunc</sub> fusion proteins for mass spectroscopy, GST-Ste20<sub>trunc</sub> fusion proteins spanning residues of 496-939 of full-length Ste20 for mass spectrometer analysis were made by transforming cells (strain BOY491 (*cln2*<sup>-</sup>) or BOY493 (*CLN2*<sup>+</sup>)) with a plasmid expressing *GAL1p::GST-Ste20*. The cells were grown in SCGal-Ura overnight, to an optical density between 0.8 and 1.0. The cells were pelleted, washed in wash buffer of 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM EDTA, and broken with glass beads in 50 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM EDTA, 0.08% Triton-X-100, plus protease inhibitors ("TNET"). Clarified extract was incubated with glutathione agarose for 1 hr at 4°C with rotation. The agarose was pelleted, washed 3 times with TNET and the protein was eluted either with an equal volume of 2X SDS-PAGE sample buffer or with 5 mM glutathione prepared in 50 mM Tris-HCl pH 8.0. Samples eluted with glutathione were concentrated using Microcon-30 microconcentrators available from Millipore Corporation, Bedford, Massachusetts.

An SDS-PAGE gel was copper stained with Bio-Rad from Hercules, CA. The Ste20<sub>trunc</sub> band (~80 kDa) was cut out, destained, washed, digested in-gel with trypsin, and the tryptic peptides extracted. HPLC separations (Michrom UMA, Michrom BioResources, Inc., Pleasanton, CA) were made with a C8 silica gel column (Inertsil C8 (150 mm x 0.7 mm, 5µm, 300Å) GL Science, Tokyo, Japan). The eluent from the HPLC column was connected directly to the electrospray ion source. Mobile phase A was acetonitrile:water (2:98 (v/v)) containing 0.1% TFA

- 26 -

and mobile phase B was acetonitrile:water (95:5) containing 0.7% TFA. A linear gradient program was run from 0 to 60% B over a period of 60 minutes (flow-rate 20  $\mu$ l/min).

Ten percent of the sample was used for MALDI-TOF-MS analysis and the other 90% for LC-MS and LC-MS/MS analysis to identify sites which were phosphorylated. LC-MS and LC-MS/MS analysis were performed with an electrospray ion trap mass spectrometer, model LCQ, available from Finnigan MAT, San Jose, CA, operated in a mode which alternated single mass spectral ("ms") scans ( $m/z$  400 – 2000) with ms/ms scans (data dependent scan mode in which the most intense ion peak in the previous ms scan was isolated and subjected to collision-induced dissociation ("CID")). The CID energy was set at 30 and the ion injection time at 100 ms. The MALDI analysis was conducted as described above in Example 2.

MALDI-TOF-MS and LC-ESI-ion trap-MS/MS analysis of unlabeled full length Ste20 as well as a truncated form spanning residues 496-939 (Ste20<sub>trunc</sub>) identified 13 sites that were phosphorylated *in vivo*: Ser<sup>418</sup> (Ser<sup>422</sup> or Thr<sup>423</sup>), Ser 502, three sites in a tryptic peptide spanning residues 506-530, Ser<sup>547</sup>, (Ser<sup>521</sup>, Thr<sup>552</sup> or Thr<sup>555</sup>), Ser<sup>562</sup>, Thr<sup>573</sup>, Ser<sup>585</sup>, Thr<sup>773</sup>, and (Ser<sup>861</sup> or Thr<sup>863</sup>). The ambiguities in the identification of certain of the sites arise because MS/MS analysis does not always provide information on each amino acid residue in the peptide sequence.

To monitor the differences in phosphorylation of STE20<sub>trunc</sub> in *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup>, two cell pools were prepared. One cell pool contained *CLN2*<sup>+</sup> in a medium containing a natural abundance of <sup>14</sup>N and the other cell pool contained *cln2*<sup>-</sup> in a medium enriched in <sup>15</sup>N. The media for each cell pool were prepared as described above in Example 2. Fractions from the two cell pools were mixed and subjected to SDS-PAGE. The band containing the mixture of labeled and unlabeled STE20<sub>trunc</sub> was excised and digested with trypsin. The resulting peptides were extracted by HPLC on a Michrom UMA instrument with a pre-column splitter and 50 mm x 0.2 mm C18 silica gel capillary column. The capillary column used was the Magicms, 200Å, 5 $\mu$ m, available from Michrom BioResources, Inc. Mobile



phase A was methanol:water (5:95) containing 1.0% acetic acid, and mobile phase B was methanol:water (85:15) containing 1.0% acetic acid. A linear gradient program was carried out from 0 to 60% of B concentration for 30 minutes. The total flow-rate was 50  $\mu$ l/min prior to pre-column splitting and  $\sim$ 3  $\mu$ l/min after splitting. The total event was directed to an electrospray ion trap mass spectrometer, which was operated in single-stage MS profile mode over the range of  $m/z$  400 – 2000 with an ion injection time of 100 ms. The mass spectrometer was a Finnigan LCQ electrospray ion trap spectrometer, available from Finnigan MAT, San Jose, California.

Fig. 11 is a schematic representation of the resulting mass spectrum. For the purpose of illustration, peptides that remain unchanged in the two cell pools are assumed to be present in equal abundance and the level of phosphorylation of peptide X is assumed to change from 30% (pool 1) to 70% (pool 2) – leading to a decrease in the measured intensity ratio of unphosphorylated peptide X and an increase for phosphorylated peptide ( $X_p$ ). Fig. 12 shows the actual changes in the ratios for the unphosphorylated peptide X and the phosphorylated peptide  $X_p$ .

Measurement of the intensity ratios of the isotopically labeled (*cln2*<sup>+</sup>) versus unlabeled (*CLN2*<sup>+</sup>) phosphopeptides showed that at least four of these sites exhibited large increases in phosphorylation in the *CLN2*<sup>+</sup> cell pool. These *Cln2*-dependent sites appear to be consensus cyclin dependent S/T-P sites, consistent with direct phosphorylation of Ste20 by *Cln2*-Cdc28.

It was found that the Ste20 peptide SKTSPIISTAHTPQQAQSPK<sup>564</sup> [Sequence I.D. No. 3] was phosphorylated at Ser<sup>547</sup>, Ser<sup>562</sup> and (Thr<sup>551</sup>, Ser<sup>552</sup>, or Thr<sup>555</sup>). The first two sites are within SP motifs (underlined) while the third site was constrained to a 5-residue stretch that contains a TP motif (underlined). Fig.13, Box A shows the actual change in the intensity ratio of the unphosphorylated peptide X as well as the change observed for the singly, doubly and triply phosphorylated Ste20 peptide obtained from *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup> cells. The intensity ratio for the unphosphorylated peptide X decreased by  $84 \pm 5$  % while that for the singly, doubly, and triply phosphorylated sites increased by  $24 \pm 12$  %,  $44 \pm 14$  %, and  $>150$  %, respectively, showing that phosphorylation is enhanced at all three sites in

the CLN2<sup>+</sup> versus the *cln2*<sup>-</sup> mutant cells. By contrast, the Ste20 phosphopeptide SLSKELNEK<sup>591</sup> [Sequence ID. No. 4] (phosphorylated on Ser<sup>585</sup>) undergoes no significant ratio change ( $9 \pm 13$  %, Fig. 13, Box B), demonstrating that phosphorylation at Ser<sup>585</sup> is not Cln2-dependent -in accord with the absence of a proline-directed kinase phosphorylation motif. Cln2-dependent phosphorylation was also observed in the peptide <sup>565</sup>APAQETVTTIPTSKPAQAR<sup>582</sup> [Sequence ID. No. 5] (Fig. 13, Box C) and to a lesser extent in the peptide <sup>772</sup>TTMVGTPYWMAPEVVSR<sup>788</sup> [Sequence ID. No. 6] (Fig. 13, Box D). Using MS/MS analysis, we found the former to be phosphorylated on Thr<sup>573</sup> (a TP motif), while the latter peptide was phosphorylated on Thr<sup>773</sup> (and not Thr<sup>777</sup>Pro<sup>778</sup>, as was previously reported in Wu et al., J. Biol. Chem. 270,15984 (1995)). These data demonstrate that the present method can precisely discern site-specific changes in the degree of phosphorylation of a protein.

Any post-translational modification of a protein which will effect the mass of the protein may be similarly compared.

The method of the present invention is applicable to a wide range of areas. The effective quantification of protein expression levels using the method of the present invention aids in understanding the interaction of gene expression with external factors in producing phenotypes. In combination with, and as an extension of genomic transcriptional expression mapping, the quantitative description of the protein phenotype assists in understanding the molecular basis of physiological and pathological processes.

In accordance with the present invention, changes in the post-translational expression of a protein or proteins with time may be studied by periodically withdrawing samples from a control cell pool and a cell pool whose metabolism has been arrested, either one of which may be isotopically labeled.

The effects of chemical compounds on the post-translational expression of proteins in one cell pool may be compared with a cell pool which has not been exposed to the compound. Such a procedure could be useful in screening drug candidates by giving an indication of side effects. For example, one cell pool can be treated with a compound which is a potential drug candidate. Comparison

with an untreated control cell pool in accordance with the present invention can indicate the changes in post-translational protein expression caused by the drug. If it is found that the drug causes a known change in the expression of protein or proteins which is known to be deleterious, that compound can be eliminated from further development prior to the commitment of significant resources. Similarly, if it is known that a change in the expression of a protein or proteins has a positive effect, a comparison of those protein levels in the treated and control pool in accordance with the method of the invention can give a preliminary determination of whether a particular compound may be a useful drug.

With a knowledge of the deleterious and positive changes in protein expression, the toxicology of pesticides, chemicals and environmental agents can also be examined in accordance with the method of the present invention. In the field of agriculture, the effects of fertilizers, pesticides and pheromones on the post-translational synthetic expression of proteins can be studied.

The effects of gene therapy can also be studied by the methods of the present invention. One or more genes in the cells in one cell pool can be inserted, replaced, modified, overexpressed or underexpressed, as is known in the art. In accordance with the present invention, a control cell pool in which the genes have not been subjected to gene therapy is provided, as well. Either one of the cell pools can be cultured in an isotopically enriched or depleted medium while the other is cultured in a medium having naturally occurring isotopic proportions. The method of the present invention enables the effects of such a genetic modification on post translational protein expression or other cellular functions to be studied.

The effects of treatments on the cell membranes themselves can also be studied in accordance with the present invention. In this case, isotopically enriched and non-isotopically enriched cell pools are combined, the cell membranes are removed from the combined cell pool in a manner known in the art, digested and subjected to mass spectroscopy.

The secreted by-products of the cell pools could also be compared by drawing samples from the media of each pool, one of which is isotopically enriched, mixing the samples, removing the cells and any other unwanted components, and

analyzing the remaining mixture of the sample media by mass spectroscopy, as described above.

The post-translational effects of hormones, infectious agents such as viruses and bacteria, carcinogens, and trauma, such as burns, can be similarly studied and quantified. Pain modulation can also be examined.

The effects of cell differentiation on post-translational protein expression can also be studied in accordance with the method of the present invention.

The applications described herein are merely examples of several of the many possible uses of the process of the present invention. While the present invention has been described with respect to quantifying post-translational changes in protein expression, changes in protein expression resulting from the effects of any modulation on translation or transcription can also be studied and quantified. In addition, the process of the present invention may be used to compare the relative quantities of any biological component which can be ionized or whose subcomponents can be ionized so that it can be analyzed by mass spectroscopy, in any type of biological matter which can be grown in an isotopically labeled medium. For example, in addition to the proteins and peptides discussed above, the biological component may be a nucleic acid, a carbohydrate, a lipid, a cofactor and post-synthetic derivatives thereof.

In addition to the biological cells discussed above, the biological matter itself may be a microbiological culture, biological tissue, an organ, an organism, a collection of organisms, a part of an organism, and a cell-free biological mimetic system, for example.

The biological component can also act as a marker for effects on biological processes. For example, if it is known that the level of a protein changes due to a change in a biological process caused by a modulation, the change in that protein level can be used to study the response of the biological process to the modulation even if the role of that protein in that process is not understood. A marker could be any biological component.

As an example of the comparison of the effects of a modulation on the characteristics of organisms, two cultures of nematodes can be grown, one in a medium containing a normal abundance of isotopes and the other in a medium isotopically enriched or depleted in at least one isotope. One of the cultures may then be modulated. The cultures can then be mixed, the nematodes broken up, such as by crushing, and a portion of the combined culture extracted and subjected to mass spectroscopy to analyze the differential effect of the modulation on the nematodes from each cell pool.

Cells, tissue, fluids or other biological matter may also be withdrawn from a human or animal subject fed isotopically enriched or depleted food for comparison with the same biological matter withdrawn from another human or animal subject fed food having normal isotopic proportions, and analyzed in accordance with the present invention. Biological matter withdrawn from different species of animals can also be compared in accordance with the present invention to study the differences in protein expression, and other differences, between the species.

Samples from the same human or animal subject can be analyzed at different points in time, as well. First, biological matter such as cells, tissues or fluids are withdrawn from the subject. Then the subject is fed food enriched in one or more isotopes. After sufficient time for metabolism, the same biological matter is withdrawn from the subject. The withdrawn materials are mixed and analyzed as described above. Samples can be withdrawn at multiple times to monitor the metabolism of the food with time, as well. Such analyses may be useful in clinical investigation and diagnosis.

As mentioned above, one of the samples of biological matter need not be cultured, grown or maintained in a medium or food having a normal abundance of isotopes. As long as the abundance of at least one isotope in one of the media or the food fed one of the animals is different from the abundance of that isotope in the other, the method of the present invention may be applied.

- 32 -

We claim:

o

1. A method for comparing the relative abundance of a protein of interest in multiple samples of biological matter comprising:

culturing a first sample of biological matter in a first medium containing a natural abundance of isotopes;

5

culturing a second sample of biological matter in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium;

modulating one of the samples;

10

combining at least a portion of each of the samples;

removing at least one protein from the combined samples;

subjecting the at least one removed protein to mass spectroscopy to develop a mass spectrum;

15

computing a ratio between the peak intensities of at least one pair of closely spaced peaks;

determining the relative abundance of the protein in each sample based on the at least one computed ratio; and

20

identifying the protein.

2. The method of claim 1, further comprising identifying the protein by electrophoresis, antibodies, bioassay or from the mass spectrum.

25

3. The method of claim 1, wherein the removing step comprises extracting proteins from the combined cell pool, the method further comprising separating the extracted proteins by a process chosen from the group consisting of one-dimensional electrophoresis, two-dimensional electrophoresis, ultra-

30

centrifugation, chromatography and affinity binding.

4. The method of claim 3, wherein the separating step comprises:

35

subjecting the extracted proteins to electrophoresis in a gel; staining the gel;

- 33 -

removing gel spots of interest; and  
destaining the gel spots of interest;  
the method further comprising digesting the proteins in the  
destained gel spots.

5                    5.     The method of claim 1, further comprising digesting the at  
least one protein of interest.

6.     The method of claim 1, further comprising computing ratios  
10 of the peak intensities of a plurality of pairs of closely spaced peaks of the mass  
spectrum.

7.     The method of claim 1, further comprising:  
15 removing a plurality of proteins of interest from the combined  
cell pool;  
digesting the plurality of proteins of interest into a plurality of  
peptides;  
20 subjecting the digested proteins to mass spectroscopy to  
develop the mass spectrum;  
selecting a plurality of pairs of closely spaced peaks on the  
mass spectrum;  
computing the ratio of the intensities of the peaks in each  
25 pair;  
determining the protein from which the pair of peaks in the  
mass spectrum are derived based on the mass spectrum; and  
determining the relative abundance of the protein in each cell  
30 pool.

8.     The method of claim 1, further comprising determining the  
relative quantity of a modified protein of interest in each cell pool.

35

- 34 -

9. The method of claim 8, further comprising determining the site of the modification on the protein.

10. The method of claim 1, wherein the modulating step comprises subjecting the biological matter in the one sample to an environmental or chemical stimulus.

11. The method of claim 1, wherein the modulating step comprises genetically manipulating the biological matter in the one sample.

12. The method of claim 11, wherein the genetic manipulation comprises performing gene therapy.

13. The method of claim 1, wherein the modulating step comprises subjecting one of the cell pools to a drug or hormone.

14. The method of claim 1, wherein the modulating step comprises subjecting the cell pool to a treatment chosen from the group consisting of a virus, a bacteria and a carcinogen.

15. The method of claim 1, wherein the protein is a marker for the effect of the modulating step on a biological process.

16. The method of claim 1, comprising culturing the second sample in a second medium in which the at least one isotope is non-radioactive.

17. The method of claim 1, comprising culturing the second sample in a second medium enriched in the at least one isotope.

18. The method of claim 1, comprising culturing the second sample in a second medium depleted in the at least one isotope.

35



- 35 -

19. The method of claim 1, wherein the biological matter is chosen from the group consisting of biological cells, biological tissue, an organ, an organism, a collection of organisms, a portion of an organism, and a cell-free biological mimetic system.

20. The method of claim 1, comprising culturing the second sample in a second medium wherein at least one isotope chosen from the group consisting of nitrogen-15, carbon-13, oxygen-17, oxygen-18, sulfur 34 and hydrogen-2, has a different abundance than the abundance of the same isotope in the first medium.

21. The method of claim 20, comprising culturing the second sample in a second medium enriched to at least about 90% in nitrogen-15.

22. A method for comparing the relative abundance of a peptide of interest derived from the same protein from multiple samples of biological matter comprising:

culturing a first cell pool in a first medium containing a natural abundance of isotopes;

culturing a second cell pool in a second medium isotopically enriched in at least one non-radioactive isotope;

modulating one of the cell pools;  
combining at least a portion of each of the cell pools;  
extracting the proteins from the combined cell pools;  
separating the extracted proteins into a plurality of proteins;  
digesting at least one protein of interest from the separated proteins into a plurality of peptides;

subjecting the digested peptides to mass spectroscopy to develop a mass spectrum; and

determining the relative abundance of the peptide of interest from each sample based on the mass spectrum.

- 36 -

23. The method of claim 22, further comprising identifying the protein from which the peptide is derived and determining the relative quantity of the protein in each cell pool.

24. The method of claim 22, further comprising determining the relative occurrence of a modification of the peptide of interest in each cell pool.

25. The method of claim 24, wherein the modification of the peptide is chosen from the group consisting of the phosphorylation, glycosylation, and acylation of the peptide, the determining step comprising determining the difference in the relative abundance of the modified peptide in each cell pool.

26. The method of claim 24, further comprising determining the level of modification of the peptide.

27. The method of claim 24, wherein the type of peptide modification is chosen from the group consisting of the phosphorylation, glycosylation, and acylation of the peptide, the determining step comprising determining the difference in the relative abundance of the modified peptide in each cell pool.

28. The method of claim 22, wherein the separating step comprises subjecting the extracted proteins to a process chosen from the group consisting of one-dimensional electrophoresis, two-dimensional electrophoresis, ultra-centrifugation, chromatography and affinity bonding.

29. The method of claim 28, wherein the separating step comprises:

subjecting the extracted proteins to electrophoresis in a gel;  
staining the gel;

removing gel spots of interest; and

destaining the gel spots of interest; and

- 37 -

o the digesting step comprises digesting the destained gel spots of interest.

30. The method of claim 29, wherein the digesting step comprises mixing the destained gel spots of interest with a proteolytic enzyme.

5

31. The method of claim 30, wherein the proteolytic enzyme is trypsin.

10 32. The method of claim 29, comprising culturing the second cell pool in a medium enriched in an isotope chosen from the group consisting of nitrogen-15, carbon-13, oxygen-17, oxygen-18, sulfur-34 and hydrogen-2.

15 33. A method for comparing the relative abundance of an ionizable component of biological matter from multiple samples, comprising:  
culturing a first sample of the biological matter in a first medium;

20 culturing a second sample of the biological matter in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium;

modulating the biological matter in one of the samples;  
combining at least a portion of each of the samples;  
25 subjecting at least a portion of the combined sample to mass spectroscopy to develop a mass spectrum; and

determining the relative abundance of at least one component of interest in each sample based on the mass spectrum.

30 34. The method of claim 33, wherein the at least one component of interest is chosen from the group consisting of a protein, a peptide, a nucleic acid, a carbohydrate, a lipid, a co-factor and post-synthetic derivatives thereof.

35

35. The method of claim 33, wherein the biological matter is chosen from the group consisting of biological cells, biological tissue, an organ, an organism, a collection of organisms, a portion of an organism, and a cell-free biological mimetic system.

36. The method of claim 33, wherein the determining step comprises computing ratios of the peak intensities of at least one pair of closely spaced peaks corresponding to the at least one component of interest of the mass spectrum.

37. The method of claim 33, further comprising culturing a third sample of the biological matter in a third medium, wherein at least one isotope has a different abundance than the abundance of the same isotope in the first and second media, and modulating the third sample by a different modulation than the modulation in the first modulating step.

38. The method of claim 33, wherein the determining step comprises computing ratios of the peak intensities of at least one pair of closely spaced peaks corresponding to the at least one component of interest of the mass spectrum.

39. The method of claim 33, comprising culturing the first sample in a first medium containing a natural abundance of isotopes.

40. A method for comparing the relative abundance of a cellular component of interest in multiple cell pools, wherein the cellular component is ionizable, the method comprising:

culturing a first cell pool in a first medium;

culturing a second cell pool in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium;

modulating one of the cell pools;

- 39 -

- combining at least a portion of each of the cell pools;  
removing at least one component of interest from the  
combined cell pool;  
subjecting the at least one removed component of interest to  
mass spectroscopy to develop a mass spectrum;  
5 computing ratios of the peak intensities of at least one pair of  
closely spaced peaks corresponding to the at least one component of interest of the  
mass spectrum; and  
identifying the component of interest.

10

41. The method of claim 40, wherein the cellular component of interest is chosen from the group consisting of a protein, a peptide, a nucleic acid, a carbohydrate, a lipid, a cofactor, and post-synthetic derivatives thereof.

15

42. The method of claim 41, wherein the component of interest is an excreted by-product of the biological matter, the method further comprising removing the biological matter from the combined cell pool and subjecting at least a portion of the medium of the combined sample to mass spectroscopy.

20

43. A method for comparing the relative abundance of biological matter from two animal subjects, wherein the biological matter can be analyzed by mass spectroscopy, comprising:

25

feeding a first animal food;

feeding a second animal food including at least one isotope having a different abundance than the same isotope in the food fed the first animal;

30

withdrawing a sample of biological matter from the first animal;

withdrawing a sample of biological matter from the second animal;

35

combining at least a portion of the withdrawn biological matter from the first and second animals;

- 40 -

subjecting the combined biological matter to mass spectroscopy to develop a mass spectrum; and  
determining the relative abundance of at least one component of interest of the combined biological matter based on the mass spectrum.

5                   44.     The method of claim 43, comprising feeding a second animal from a species different than the species of the first animal.

10                   45.     The method of claim 43, comprising providing first and second human subjects.

15                   46.     A method for comparing the relative abundance of biological matter from a single animal subject at different times, wherein the biological matter can be analyzed by mass spectroscopy, comprising:

withdrawing a first sample of biological matter from the subject;

20                   feeding the subject food including at least one isotope having a different abundance than the same isotope in the food fed the first animal, after withdrawing the first sample;

withdrawing a sample of biological matter from the subject after feeding;

25                   mixing at least portions of the first and second samples;  
subjecting the combined samples to mass spectroscopy to develop a mass spectrum; and

determining the relative abundance of at least one component of interest of the combined biological matter based on the mass spectrum.

30                   47.     The method of claim 46, comprising providing a human subject.

35

- 41 -

48. The method of claim 46, further comprising withdrawing at least one subsequent sample of biological matter from the subject a period of time after withdrawing the second sample, mixing at least a portion of the at least one subsequent sample with a portion of the first sample and then performing the subjecting and determining steps to monitor the metabolism of the food with time.

5

10

15

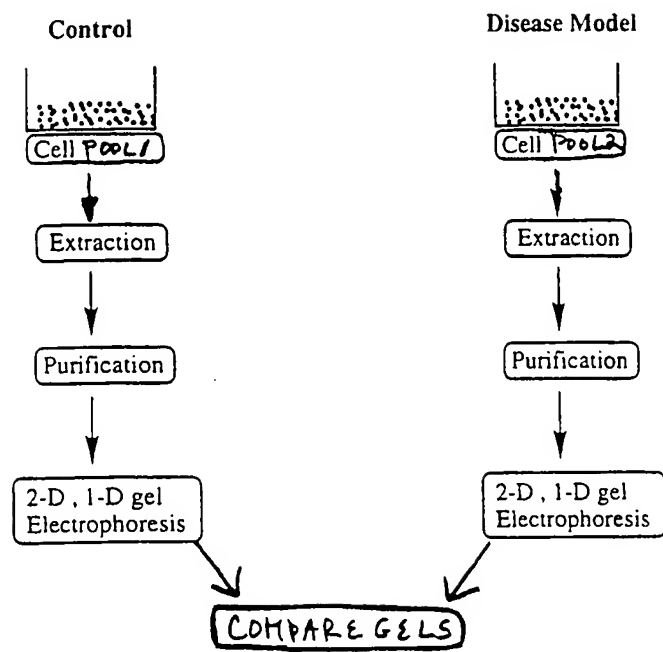
20

25

30

35

1/11

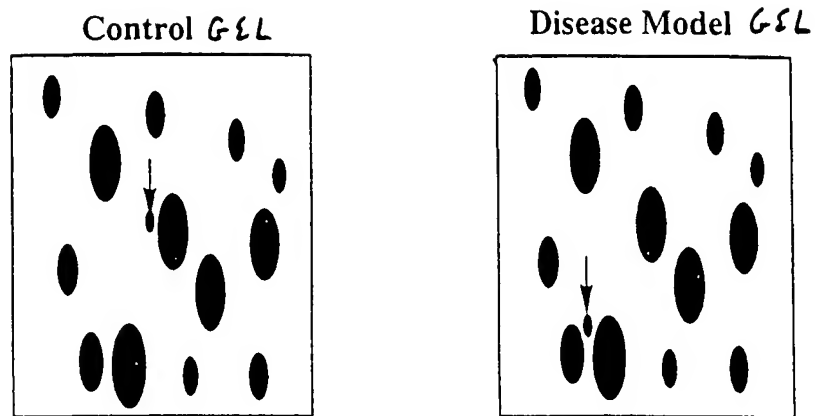


PRIOR ART

FIG. 1



2/11



PRIOR ART

FIG. 2

3/11

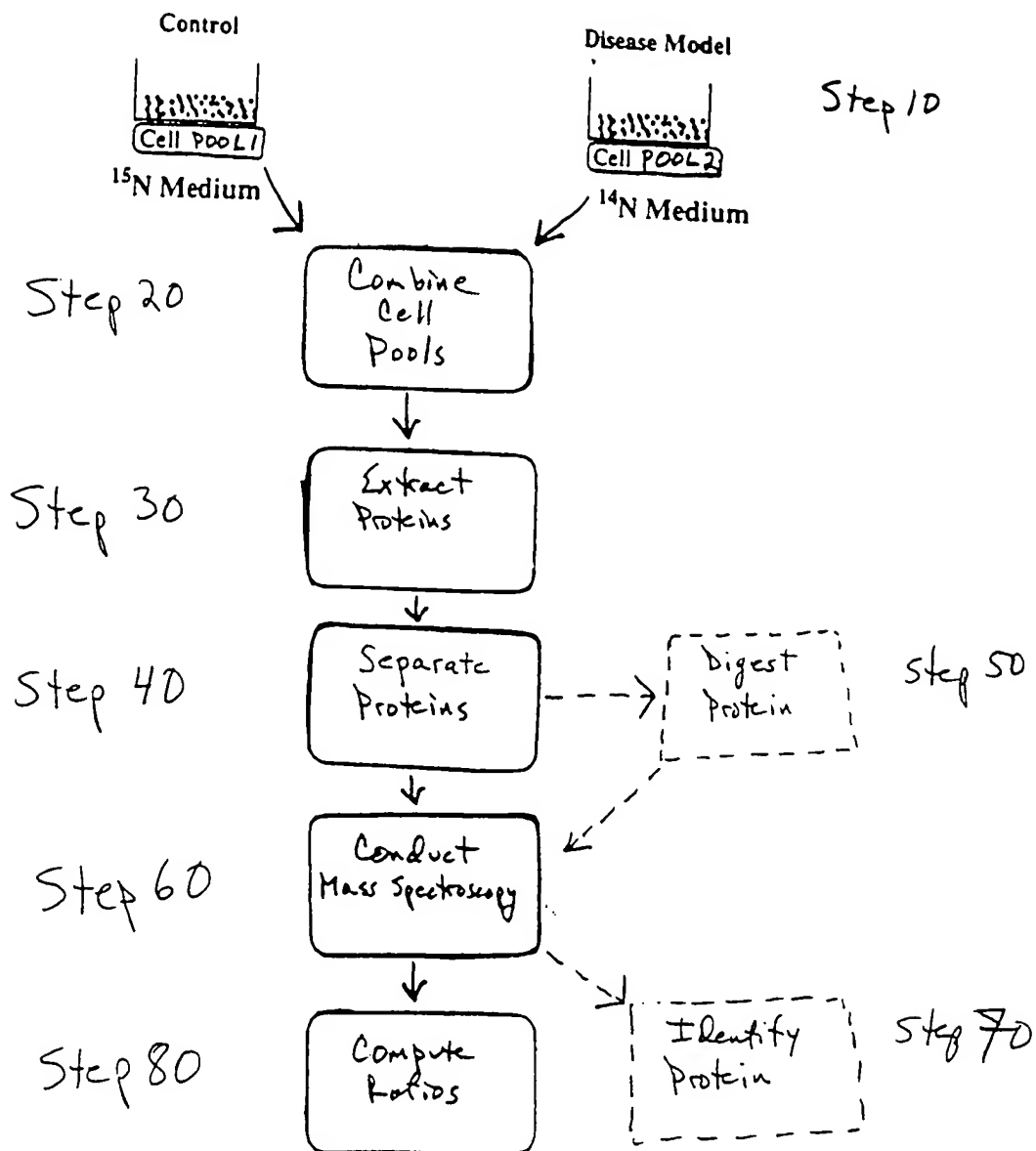


FIG. 3

4/11

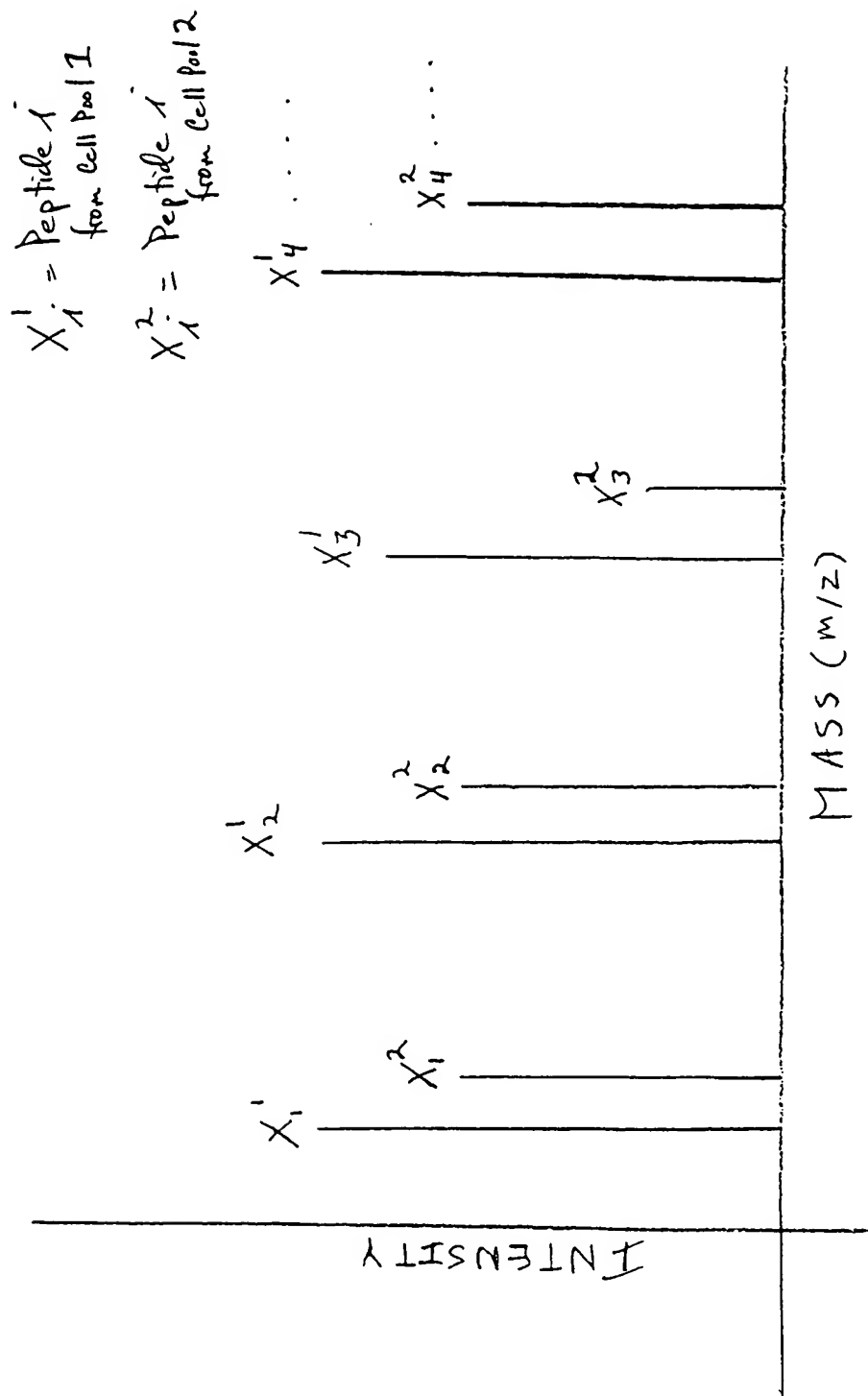


Fig. 4

5/11

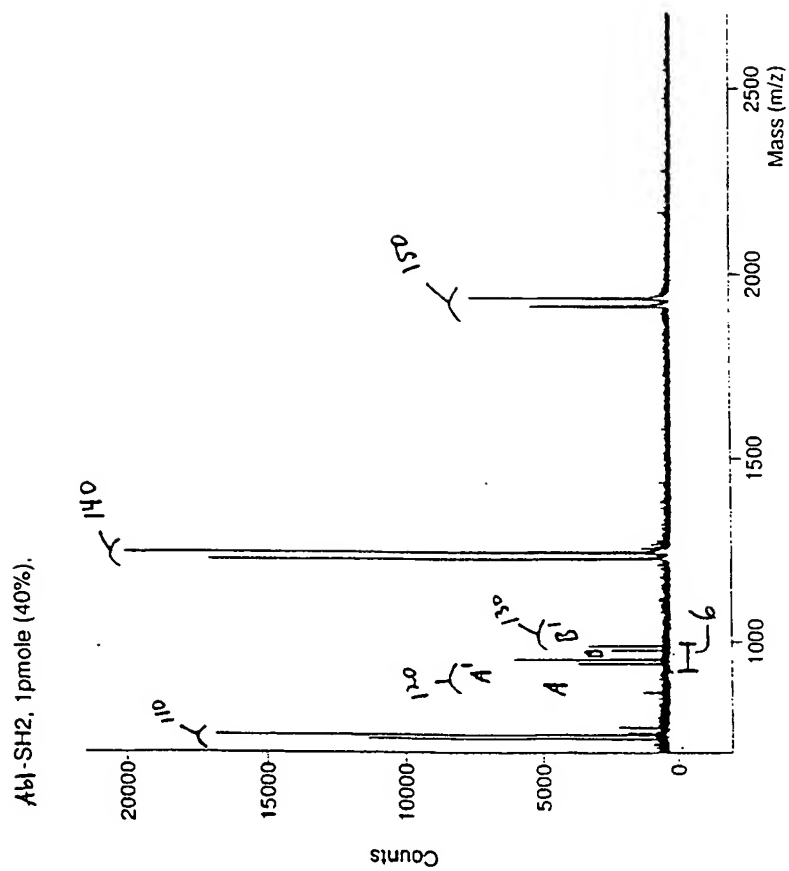


FIG. 5

6/11

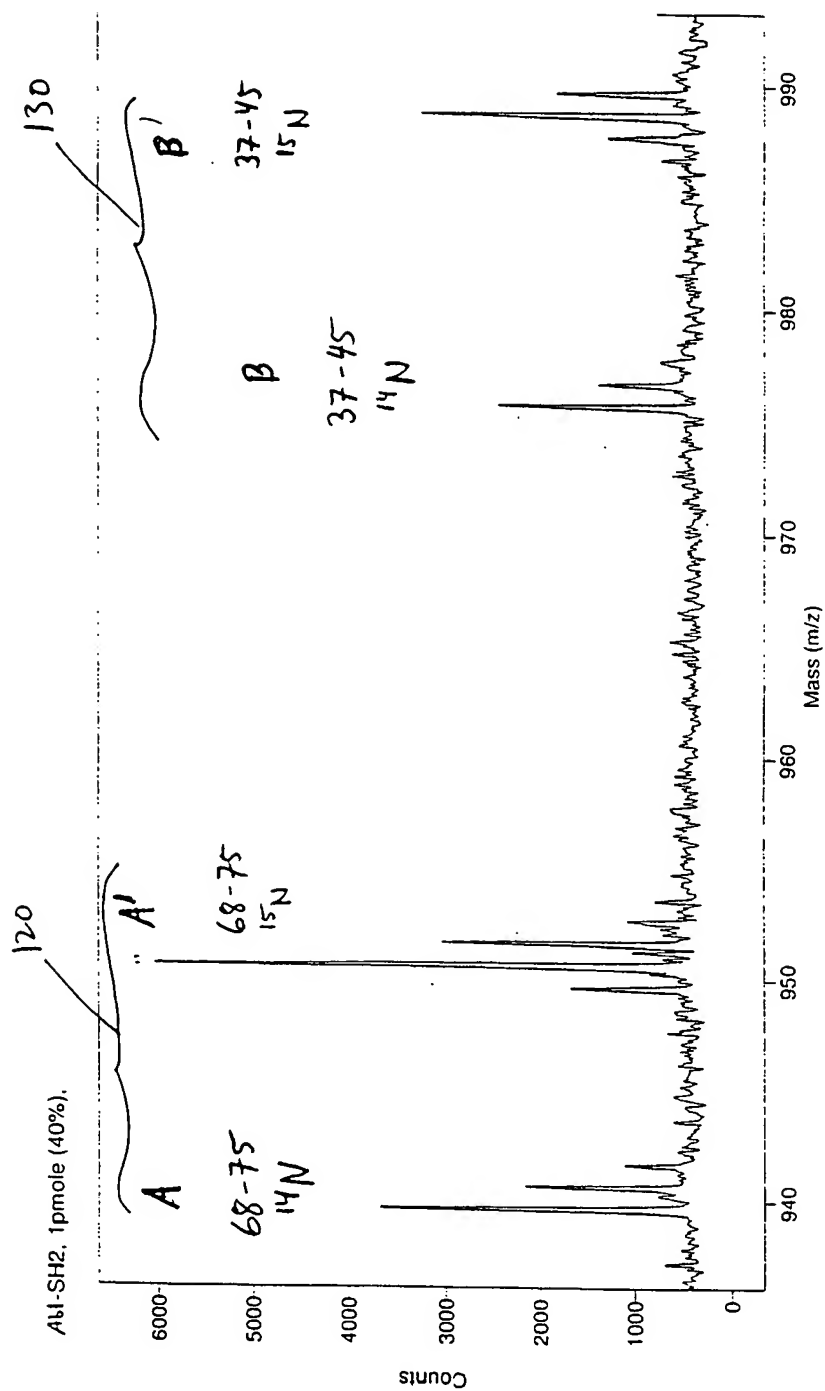


FIG. 6

7/11

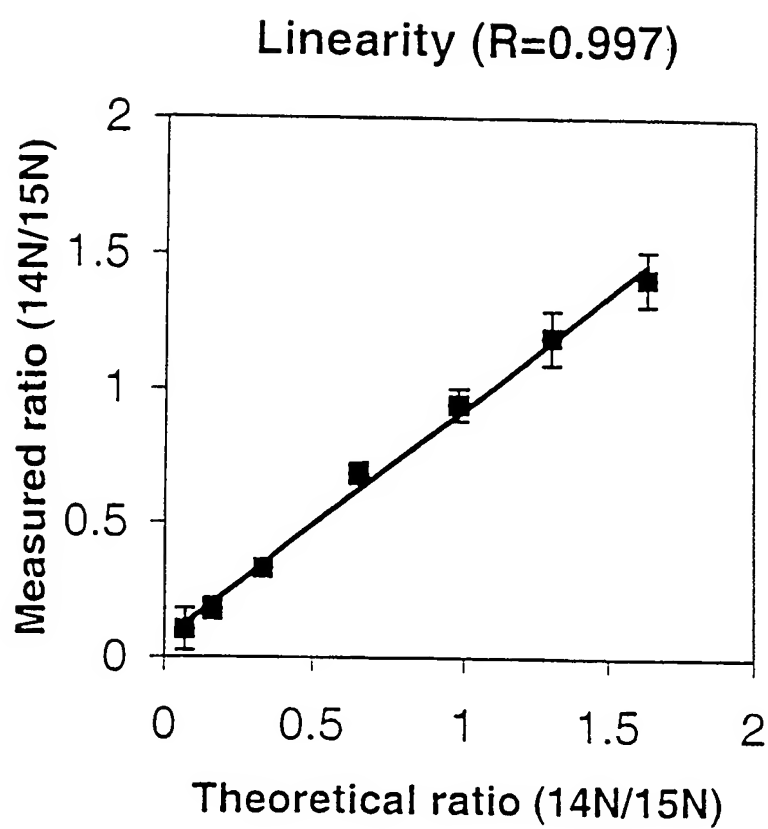


Fig. 7

8/11

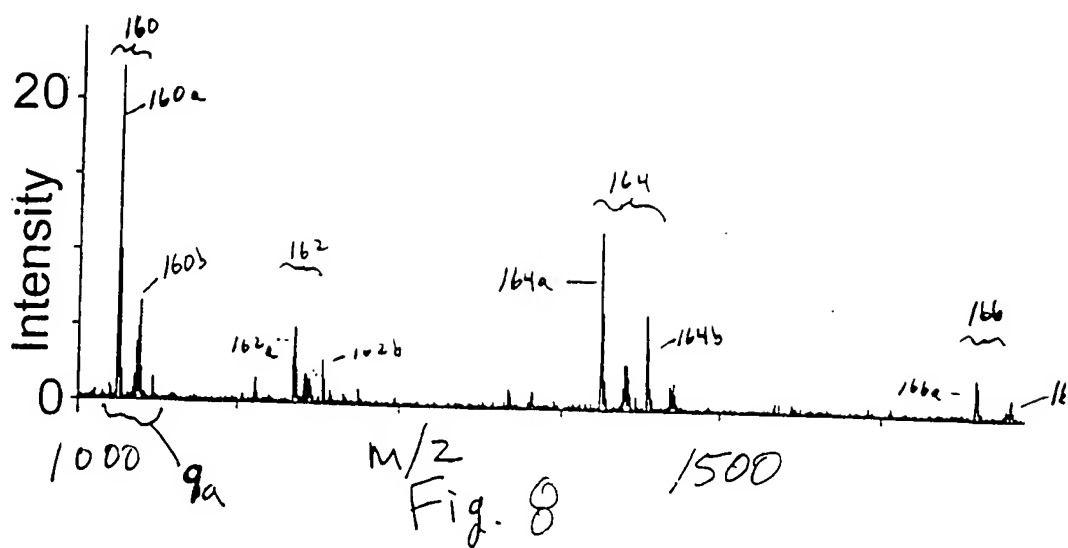


Fig. 9a

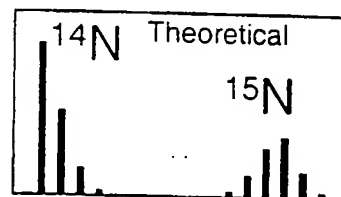
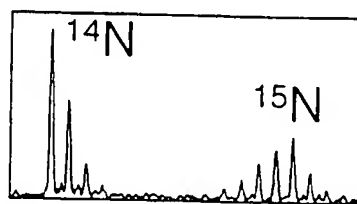


Fig. 9b

9/11

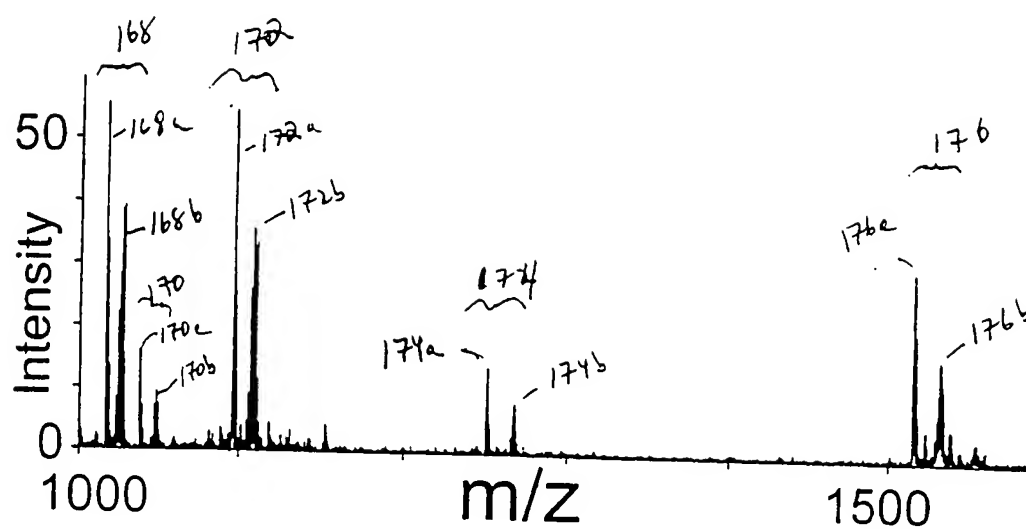
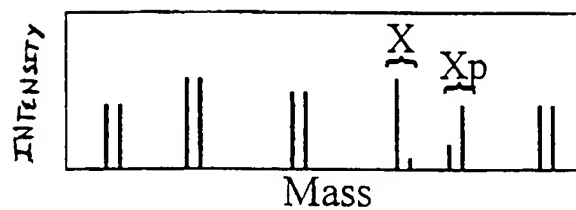


Fig. 10



10/11



X: Unphosphorylated peptide X

Xp: Phosphorylated peptide X

Fig. 11

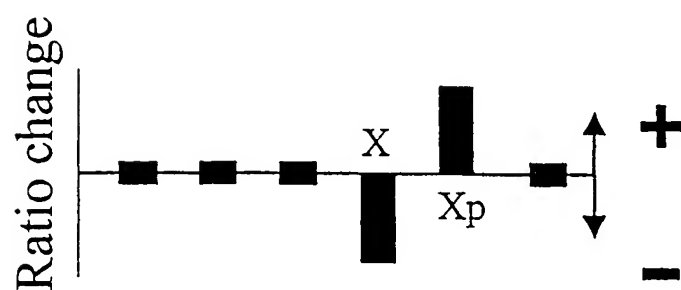


Fig. 12

11/11

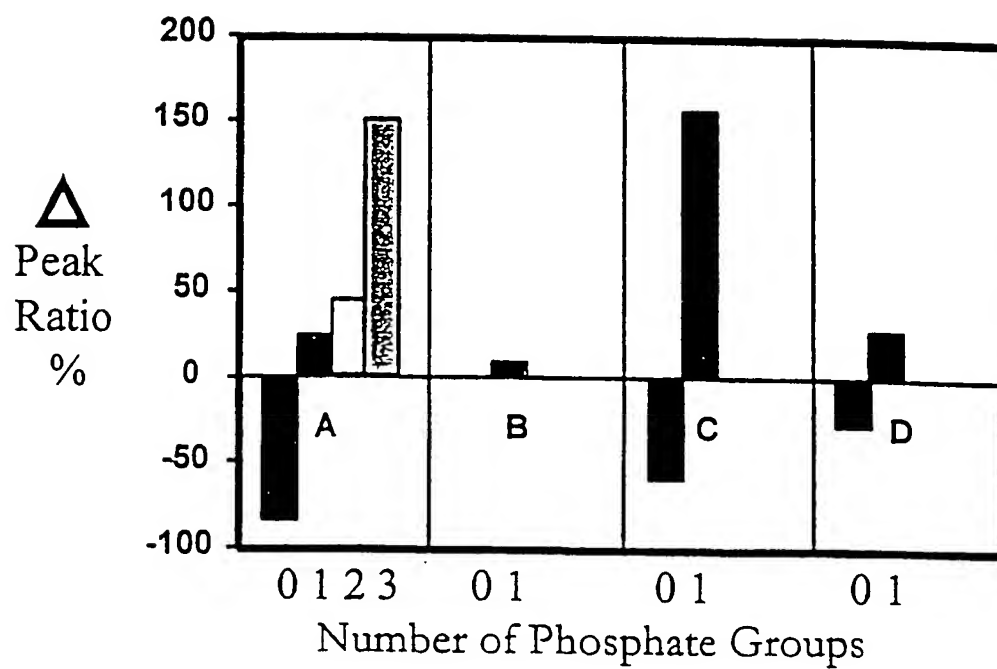


Fig. 13

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US00/12026

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G01N 33/00, 24/00

US CL :435/71.1; 436/86,173

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/71.1; 436/86,173

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
EAST

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,800,979 A (KOLHOUSE et al) 01 September 1998, see entire document.	1-48
A	US 5,572,024 A (GRAY et al) 05 November 1996, see entire document.	1-48

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

14 AUGUST 2000

Date of mailing of the international search report

28 AUG 2000

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

*L. Blaine Lankford*  
L. BLAINE LANKFORD

Telephone No. (703) 308-0196

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
9 November 2000 (09.11.2000)

PCT

(10) International Publication Number  
**WO 00/67017 A1**

- (51) International Patent Classification<sup>7</sup>: **G01N 33/00**, 24/00 (74) Agents: **SKLAR, Brandon, N. et al.**: Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).
- (21) International Application Number: PCT/US00/12026 (81) Designated States (*national*): CA, JP.
- (22) International Filing Date: 3 May 2000 (03.05.2000) (84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
09/304,799 4 May 1999 (04.05.1999) US
- (71) Applicant: **THE ROCKEFELLER UNIVERSITY** [US/US]; 1230 York Avenue, New York, NY 10021-6399 (US). (48) Date of publication of this corrected version: 13 June 2002
- (72) Inventors: **CHAIT, Brian, T.**; 500 East 63rd Street, Apt. 20D, New York, NY 10021 (US). **COWBURN, David**; 215 East Dudley Avenue, Westfield, NJ 07090 (US). **ODA, Yoshi**; #102 4-8-1 Ninomiya, Tsukuba, Ibaraki 305-0051 (JP). (15) Information about Correction: see PCT Gazette No. 24/2002 of 13 June 2002, Section II
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: METHOD FOR THE COMPARATIVE QUANTITATIVE ANALYSIS OF PROTEINS AND OTHER BIOLOGICAL MATERIAL BY ISOTOPIC LABELING AND MASS SPECTROSCOPY

(57) Abstract: The present invention is a method for accurately comparing the levels of cellular components, such as proteins, present in samples which differ in some respect from each other using mass spectroscopy and isotopic labeling. A first sample of biological matter, such as cells, is cultured in a first medium and a second sample of the same biological matter is cultured in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium. One of the samples is modulated, such as by treatment with a bacteria, a virus, a drug, hormone, a chemical or an environmental stimulus. The samples are combined and at least one protein is removed. The removed protein is subjected to mass spectroscopy to develop a mass spectrum. A ratio is computed between the peak intensities of at least one closely spaced pair of peaks to determine the relative abundance of the protein in each sample. The protein is identified by the mass spectrum or through other techniques known in the art. Modifications to the proteins, such as the phosphorylation of the protein, and the site of the modification may also be determined through the process of the present invention. The method is applicable to the components of any type of biological matter which are ionizable and may therefore be analyzed by mass spectroscopy.

WO 00/67017 A1

**METHOD FOR THE  
COMPARATIVE QUANTITATIVE ANALYSIS OF  
PROTEINS AND OTHER BIOLOGICAL MATERIAL BY  
ISOTOPIC LABELING AND MASS SPECTROSCOPY**

**Statement as to Rights to Inventions Made Under  
Federally-Sponsored Research and Development**

The U.S. Government has certain rights to the invention, by virtue of its partial support of research under National Institute of Health Grant Nos. RR00862 (BTC), GM-47021 and GM49716 (FRC).

**Field of the Invention**

The invention relates to the analysis of biological matter and, more particularly, to the comparison of isotopically labeled components of biological matter from one sample with the same, unlabeled components of biological matter from another sample, through mass spectroscopy. The method is particularly suited for quantifying differences in protein expression or modification in two cell populations or pools, one of which is subjected to environmental, genetic or chemical modulation.

**Background of the Invention**

Many biological processes in living cells are controlled by alterations in the levels or states of certain key proteins. Measuring the levels of the various proteins that affect (or are affected by) the process is therefore important for gaining an understanding of the biological process. For example, a given hormone may, through a signaling cascade, activate certain key transcription factors which in turn induce the expression of a number of proteins with distinct activities. Comparison of the levels of the proteins in the cell prior to and after induction can indicate which gene products are being up regulated and/or down regulated by the action of the hormone. As a second example, comparison of the total complement of proteins from an organism (i.e., the proteome) prior to and after infection with a virus can show which proteins are down and/or up regulated by the infection. Such an analysis can provide important information about the mechanism by which the virus

- 2 -

subverts its host cell, thereby aiding in the development of anti-viral drug strategies. Similarly, comparison of some or all of the proteins of the proteome before and after treatment with a drug can indicate the mechanism of action of the drug, as well as its potential effectiveness and toxicity. As another example, measurement of the state of phosphorylation of protein members of an intracellular cascade involved in turning on and off a given biological process can provide information about the control of the signaling pathway.

A facile method for accurately comparing the levels of proteins and other cellular components and biological materials as a function of time or as the result of particular treatment, such as a hormone, a drug, or a virus, as mentioned above, or an environmental stimulus, such as a temperature change, is needed. It is also necessary to assay these protein levels with high accuracy because small changes in the levels of certain key proteins may, through a complex cascade of molecular events, produce large changes in the biological system.

Two-dimensional electrophoresis has been used to compare proteins from different cell cultures or hosts subjected to differing conditions. See, for example, Anderson, N.G., et al., "Simultaneous Measurement of Hundreds of Liver Proteins: Application in Assessment of Liver Function," Toxicologic Pathology, 1996, Vol. 24, No. 1, pp. 72-76; Anderson, N.G., et al., "Twenty years of two-dimensional electrophoresis: Past, present and future," Electrophoresis, 1996, Vol. 17, pp. 443-453; Anderson, N.G., "Covalent Protein Modifications and Gene Expression Changes in Rodent Liver Following Administration of Methypyrilene: A Study Using Two-Dimensional Electrophoresis," Fundamental and Applied Toxicology, 1992, Vol. 18, pp. 570-580; and Anderson, N.G., et al., "Global Approaches to Quantitative Analysis of Gene-Expression Patterns Observed by use of Two-Dimensional Gel Electrophoresis," Clin. Chem. 1984, Vol. 30, No. 12, pp. 2031-2036.

Fig. 1 is a schematic representation of the processes described in these articles. Proteins from a control cell culture are extracted, purified and separated by one- and two-dimensional electrophoresis. Proteins from another, parallel cell culture, which may include cells exposed to drugs, carcinogens or other

such treatments directly or through a host, are also extracted, purified and separated by one- and two-dimensional electrophoresis. Fig. 2 shows exemplary electrophoretic gel samples from each cell culture. Spots at different locations in each gel sample may indicate the presence of different proteins or changes in the proteins in the control versus the treated cells. Spots of different sizes may indicate a change in the quantity of the protein in the treated cells. The gels may be analyzed visually or by labeled maps, bargraphs or numerical tables. See, Anderson, "Twenty years of two-dimensional electrophoresis....", at p. 450. Computer generated arrowplots, which indicate the magnitude and polarity of changes in spots between gel samples of a control and treated cell pool, superimposed on a gel sample of a control cell pool, have also been used. Id. Instead of a control cell sample, the gel sample including the proteins from the treated cells may be compared to a master gel pattern from a library of gel patterns.

Using prior art methods, hundreds of gel samples and hundreds of thousands of protein abundance measurements may be required in a typical study. Id. It is also difficult to maintain the reproducibility of the extraction and purification procedures in each cell sample. Extraction and purification results must be normalized. Precise, accurate and reproducible quantification of the changes between cell pools is also difficult. If a gel spot includes more than one protein, the discrete proteins frequently cannot be identified. Thus, a more practical method of comparing proteins in different cell pools is needed.

Mass spectroscopy is a highly accurate analytical tool for determining molecular weights and identifying chemical structures. Proteins and peptides have been studied by matrix-assisted laser desorption mass spectroscopy and electrospray ionization mass spectroscopy. See, for example, Chait, Brian T. and Kent, Stephen B.H., "Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins", Science, Vol. 257, September 25, 1992, pp. 1885-1894, which is incorporated by reference herein. Matrix-assisted laser desorption time-of-flight mass spectrometers are described in U.S. Patent Nos. 5,045,694 and 5,453,247, to Beavis, et al., which are assigned to the assignee of the present invention and incorporated by reference herein. Electrospray ionization

mass spectrometers are described in U.S. Patent No. 5,245,186 to Chait et al., and U.S. Patent No. 4,977,320 to Chowdhury et al., for example, which are also assigned to the assignee of the present invention and incorporated by reference herein. Prior to analysis, the proteins are typically separated by one- or two-dimensional electrophoresis and then digested by an appropriate enzyme. The resulting peptides are then subjected to mass spectroscopy by any of the types of mass spectrometers identified above.

However, quantitative comparisons among proteins within a sample or between samples may be compromised by a number of parameters, such as the ionization efficiency of the mass spectrometer for a particular peptide or protein, the extraction efficiency from electrophoretic gels for a particular peptide and the digestion efficiency of an enzyme at different cleavage sites.

Isotopic labeling by stable or radioactive isotopes has been used to study many aspects of human, animal and plant metabolism. For example, isotopic labeling has been used to study metabolic turnover rates and biosynthesis of proteins and nucleic acids. Microorganisms, organs and tissue extracts, for example, may also be studied through isotopic labeling. The presence of radioactive isotopes in a sample of biological material may be detected by scintillation counters, or autoradiography, for example. However, the use of radioactive isotopes pose hazards to those conducting the experiments and require the use of protective measures, which may be cumbersome and expensive. To avoid this problem, in U.S. Patent No. 5,366,721, a long-lived radioisotope, such as carbon-14, is administered to a biological host. A reacted fraction is isolated from the host and the radioisotope concentration is measured by mass spectroscopy. See also DeLeecher, A.P. et al., "Applications of isotope dilution - mass spectrometry in clinical chemistry, pharmacokinetics, and toxicology," Mass Spectroscopy Reviews, 1992 11, 249-307; Grostic, M.F. et al., "Mass-Spectral Studies Employing Stable Isotopes in Chemistry and Biology," appearing in Mass Spectroscopy: Techniques and Applications, edited by Mike, G.W.A., Wiley-Interscience (1971), pp. 217-287.



### Summary of the Invention

The present invention is a method for accurately comparing the levels of ionizable components of biological matter, wherein the biological matter differs in some respect from each other, using mass spectroscopy and isotopic labeling.

In one embodiment of the present invention, a method for comparing the relative abundance of a protein of interest in multiple samples of biological matter is disclosed, wherein one of the samples has been modulated by exposure to a treatment, such as a bacteria, virus, drug or hormone, or a stimulus, such as a chemical or environmental stimulus. A first sample of the biological matter is cultured in a first medium containing a natural abundance of isotopes and a second sample of the biological matter is cultured in a second medium containing more or less than the natural abundance of one or more isotopes. One of the samples is modulated, at least portions of the samples are combined and at least one protein is removed from the combined sample.

The removed protein, which may or may not be digested into peptides, is subjected to mass spectroscopy to develop a mass spectrum. The difference in the mass of the isotope in each cell pool results in two distinct, closely spaced peaks for each protein or peptide in the mass spectrum. One peak corresponds to a protein or peptide from a protein from the cell pool with the normal abundance of isotopes. The other peak corresponds to a protein or peptide from the cell pool enriched in one or more of the isotopes. A ratio is computed between the peak intensities of at least one pair of peaks in the mass spectrum. The relative abundance of the protein in each sample may be determined based on the computed ratio. The protein may be identified by the mass-to-charge ratios of the peaks in the mass spectrum, as well as by other means known in the art.

In addition, modifications, such as phosphorylation, glycosylation or acylation, at specific sites on individual proteins may be detected and quantified through mass spectroscopy in accordance with the present invention.

The first sample need not contain a natural abundance of isotopes, as long as at least one isotope in the second sample of biological matter has a different abundance than the abundance of the same isotope in the first sample. Preferably,

- 6 -

- o the isotope which is enriched or depleted is a non-radioactive isotope of nitrogen, oxygen, carbon and/or sulfur. Hydrogen may be used, as well. Radioactive isotopes may also be used.

The effects of two or more modulations can be simultaneously analyzed by preparing additional samples with media containing an isotope with a different abundance than the abundance of the isotope in the other samples, and modulating the additional samples.

Up to the point of the mass spectroscopy, none of the steps of the process discriminates between a protein that contains the natural abundance of isotopes from the same protein from the enriched sample. Thus, the ratios of the original amounts of proteins from the two samples are maintained, normalizing for differences between extraction and separation of the proteins in the samples.

The method is applicable to the components of any type of biological matter which can be ionized and therefore may be analyzed by mass spectroscopy. For example, the component may be a protein, a peptide, a carbohydrate, a lipid, a cofactor and post-synthetic derivatives thereof. The biological matter may be a culture of biological cells, a microbiological culture, biological tissue, an organ, an organism, a collection of organisms, a part of an organism, and a cell-free biological mimetic system, for example.

### **Brief Description of the Figures**

Fig. 1 is a flow chart of a prior art process for comparing proteins from two cell pools, one of which is diseased;

Fig. 2 shows two exemplary electrophoretic gel samples from the process of Fig. 1;

Fig. 3 is a flow chart of one embodiment of the method of the present invention;

Fig. 4 is an exemplary mass spectrum resulting from a hypothetical experiment conducted in accordance with the method of Fig. 3;

Fig. 5 is a portion of a mass spectrum of Abl-SH2;

Fig. 6 is an enlarged view of the region 6 of the mass spectrum of Fig. 5;

Fig. 7 is a graph of measured ratios between the peak intensities corresponding to peptides of Abl-SH2 from an unlabeled (nitrogen-14) sample and the peak intensities corresponding to a labeled (nitrogen-15) sample, versus the expected ratios, based on seven spectral measurements of each of the samples of Table II.

Fig. 8 is a portion of a mass spectrum of the peptide of the protein elongation factor 1- $\alpha$  from a combined pool of labeled and unlabeled *Saccharomyces cerevisiae* which differed only in their ability to express the cyclin *CLN2*;

Fig. 9a is an enlarged view of region 9a in the mass spectrum of Fig. 8, showing a pair of peaks from a single peptide;

Fig. 9b shows the theoretical isotope distributions of nitrogen-14 and nitrogen-15;

Fig. 10 is a portion of a mass spectrum of the peptides of the protein triosephosphate isomerase from the combined cell pool described above with respect to Fig. 8;

Fig. 11 is a schematic representation of a mass spectrum, indicating a plurality of pairs of peaks, including a pair of peaks corresponding to a unphosphorylated peptide (X) and a pair of peaks corresponding to the same peptide, which has been phosphorylated ( $X_p$ );

Fig. 12 is bar graph showing the changes in the ratios of the unphosphorylated peptide X and the phosphorylated peptide  $X_p$ ; and

Fig. 13 shows the experimentally observed change of the intensity ratio of the unphosphorylated peptide as well as the change observed for the singly, doubly and triply phosphorylated Ste20 peptide obtained from *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup> cells.

### **Detailed Description of the Invention**

Fig. 3 is a flow chart of one embodiment of the process of the present invention for comparing the quantities of proteins in different cell samples. Two cell pools are prepared. (Step 10). One of the cell pools, here Cell Pool 1, contains a cell culture grown or maintained in a medium containing a natural abundance of

- 8 -

isotopes. For example, the medium contains 99.6% nitrogen-14 and 0.4% nitrogen-15, which are the naturally occurring abundances of those isotopes of nitrogen. The isotopes of oxygen, carbon, sulfur and other elements are also present in their naturally occurring abundances.

The other cell pool, here Cell Pool 2, contains a cell culture grown or maintained in a medium in which one or more isotopes of nitrogen, carbon, oxygen or sulfur, for example, is not present in a natural abundance. For example, the second medium may be isotopically enriched in one or more of the following isotopes: nitrogen-15, carbon -13, oxygen -17, oxygen -18, and sulfur -34.

Enrichment of hydrogen-2 (deuterium) can also be used, however, the enzymatic effects of hydrogen substitution is greater than that of  $^{15}\text{N}$  and can therefore change the biological process.

High enrichment is preferred. The particular level of enrichment may depend on the isotope chosen. For example, a medium enriched in nitrogen -15 ( $^{15}\text{N}$ ) to between 90% - 100% is preferred, with 100%  $^{15}\text{N}$  enrichment most preferred. Because hydrogen -2 can be toxic, less than 100% hydrogen -2 is preferred. Isotopic depletion may also be used wherein less than the naturally occurring abundance of an isotope is provided in the medium. Non-radioactive isotopes are preferred. Radioactive isotopes may also be used, but are not preferred because of the difficulties the use of such isotopes present, as discussed above.

The first cell pool need not contain a natural abundance of isotopes, as long as at least one isotope in the second cell pool has a different abundance than the abundance of the same isotope in the first cell pool.

It is preferred that the media the cell pools are grown in be identical, except for the presence of the different isotope or isotopes. Bio-Express from Cambridge Isotope Laboratories, Inc., Andover, Massachusetts, is one suitable  $^{15}\text{N}$  enriched media. Alternatively, the cells may be grown in the same medium and the labeling isotope may be added directly to one culture of cells.

Either of the cell pools, here Cell Pool 2, is modulated by a bacteria, a virus, a chemical, a drug, a hormone, or an environmental change, such as a

temperature change, for example. Other treatments or stimulus may be provided, as well. The other cell pool, here Cell Pool 1, acts as a control.

All or portions of the cell pools are then combined. (Step 20). The proteins are extracted from the combined cell pool in a manner known in the art. (Step 30). For example, the cell membranes may be digested or disrupted by standard methods, such as detergents or homogenization in an isotonic sucrose solution. The proteins are then extracted from the combined cell pools by ultra-centrifugation, or other known techniques. For example, antibodies may be used to immunoprecipitate certain proteins or complexes of proteins, as well. The particular method used may be dependent on the particular proteins of interest, as is known in the art.

The mixture of proteins is then separated into the individual proteins or small groups of proteins, also by known techniques, such as one- and/or two-dimensional electrophoresis, ultra-centrifugation, chromatography or affinity binding. (Step 40). Two-dimensional sodium dodecylsulfate-polyacrylamide gel electrophoresis ("SDS-PAGE"), may be used, for example. If an individual protein is extracted from the combined cell pool, such as by use of an antibody, the separation step 40 may not be necessary.

The separated proteins are then preferably digested into peptides. (Step 50). Preferably, the proteins are digested by a proteolytic enzyme. Trypsin is preferred because it cleaves precisely at the sites of lysine and arginine, yielding doubly-charged peptides which typically have a length of from about 5 to 50 amino acids and a molecular weight of between about 700-5,000. Such peptides are particularly appropriate for analysis by mass spectroscopy, especially by electrospray ionization mass spectroscopy. Other site specific proteolytic enzymes which may be used include Ly-C, Asp-N and Glu-C, for example. Pepsin, subtilisin and proteinase 1c are low specificity enzymes which may also be used. Chemical reagents may also be used to digest the proteins. For example, cyanogen bromide may be used to cut a protein into peptides at the site of methionine. BNPS-skatole may be used to cleave at the site of tryptophan. Acid hydrolysis may also be used.

o The proteins or digested proteins are then subjected to mass spectroscopy. (Step 60). Any mass spectrometer may be used to analyze the peptides or proteins. For example, the mass spectrometer may be a Matrix-Assisted Laser Desorption/Ionization ("MALDI") Time-of-Flight ("TOF") Mass Spectrometer, available from PerSeptive Biosystems, Framingham, Massachusetts; 5 an Electrospray Ionization ("ESI") ion trap mass spectrometer, available from Finnigan MAT, San Jose, California; or an ESI quadrupole mass spectrometer, available from Finnigan MAT or the Perkin-Elmer Corporation, Foster City, California.

10 A simple mixture of from 1 to about 5 digested proteins can be analyzed by single-stage mass spectroscopy with any of the mass spectrometers discussed above. Mixtures of greater than six digested proteins are preferably analyzed by a two-stage tandem mass spectroscopy procedure involving collision 15 produced dissociation ("CID"), as is known in the art.

While preferred, the digestion step 50 is not required. One or several whole proteins can also be subjected to mass spectroscopy, avoiding the need for digesting the proteins into peptides, as is known in the art. Single-stage mass 20 spectroscopy may be used to analyze mixtures of large numbers of whole proteins simultaneously.

The protein or proteins subjected to the mass spectroscopy are also preferably identified. (Step 70). The identification step 70 can take place at any time after separation or extraction of a single protein. Protein identification 25 software which uses algorithms to compare the mass spectrum with a database of proteins are available. One such algorithm, ProFound, uses a Bayesian algorithm to search protein or DNA databases to identify the optimum match between the experimental data and the protein in the database. W. Zhang, B.T. Chait, 30 "Proceedings of the 43<sup>rd</sup> ASMS Conference on Mass Spectroscopy and Allied Topics," Atlanta, Georgia (1995) p. 643. ProFound may be accessed on the World-Wide Web at <<http://prowl.rockefeller.edu>> and <<http://www.proteometrics.com>>. ProFound accesses the non-redundant database (NR). Alternative algorithms for 35 protein identification include: Mass Search (<http://cbrg.inf.ethz.ch/subsection>

3\_3.html); MOWSE (<http://www.seqnet.dl.ac.uk/mows.html>); MSFIT  
(<http://prospector.ucsf.edu/ucshtml/msfit.htm>); Peptide Mass Search  
([http://www.mdc-berlin.de/~emu/peptide\\_mass.html](http://www.mdc-berlin.de/~emu/peptide_mass.html)); and Peptide Search  
([http://www.mann.embl\\_heidelberg.de/services/peptide\\_search/fr\\_peptide  
searchform.html](http://www.mann.embl_heidelberg.de/services/peptide_search/fr_peptide_searchform.html)). See also, James, Peter, "Protein identification in the post-genome  
era: the rapid rise of proteomics", *Q. Rev. Biophysics*, Vol. 30, No. 4, pp. 279-331  
(1997), which is incorporated by reference, herein.

The protein can also be identified by electrophoresis, antibodies  
Edman sequencing or bioassay, or by other methods conventionally used in the art,  
after separation of the proteins in step 40.

The ratios of the peak intensities of each pair of peaks are then  
computed. (Step 80). The ratios give a measure of the relative amount of that  
peptide in each cell pool, as discussed further, below. The peak intensities are  
calculated in a conventional manner.

Fig. 4 is an exemplary mass spectrum of four peptides from a single  
protein resulting from a hypothetical experiment conducted in accordance with the  
method of Fig. 3.

Because of the difference between the masses of the peptides or  
proteins and resulting peptides from cells grown in the non-labeled medium and  
those grown in the labeled, isotopically enriched medium, the results of the mass  
spectroscopy will generally be a plurality of pairs of closely spaced peaks, each peak  
being at a different  $m/z$  ratio. Since the enriched isotope is typically heavier than  
the most abundant naturally occurring isotope ( $^{15}\text{N}$  versus  $^{14}\text{N}$ , for example), the  
peak at the higher  $m/z$  ratio is generally indicative of the relative abundance of the  
peptide from a labeled protein grown in the medium enriched in one or more  
isotopes. The peak at the lower  $m/z$  ratio is generally indicative of the relative  
abundance of the peptide from an unlabeled protein in the medium containing a  
normal abundance of isotopes.

Table I, below, lists four ratios of the peak intensities of isotopically  
labeled peptides of Cell Pool 2 and non-isotopically labeled peptides of the control  
Cell Pool 1 ( $X_i^2/X_i^1$ ) based on the exemplary mass spectrum of Fig. 4.

TABLE I

	Peptide No.	Intensity Ratio
	1	$0.70 \pm .02$
	2	$0.69 \pm .02$
	3	$0.30 \pm .02$
5	4	$0.71 \pm .02$

Since the number of cells in one cell pool may differ from the number of cells in the other cell pool, for any given pair of peaks, the intensity of a peak corresponding to a peptide from one cell pool may differ from the intensity of the peak corresponding to the same peptide from the other cell pool. The ratios between most of the pairs of peaks (which are indicative of peptides derived from proteins unaffected by a given treatment or stimulus), will generally be the same. A deviation from the regularly-observed ratio indicates a difference in the relative quantity of a peptide, and hence a protein, between the two cell pools which may be caused by the modulation to which one of the cell pools has been subjected. The difference can be quantified in accordance with the present invention.

Because the cell pools are combined, other sources of differences in the intensities of the peaks, such as variations in the extraction efficiency of a particular protein from the cell pool, the subsequent extraction efficiency of a protein from the gel, the digestion efficiency of the enzyme used (if any), the ionization efficiency of the mass spectrometer for a particular peptide, and other such factors, affect both cell pools equally. These factors should not, therefore, affect the observed ratios. Analyzing the ratio of the pair of peaks compensates for differences in mass intensities resulting from differences in the ionization efficiency of the mass spectrometer for a particular peptide. Isotopically labeling one of the two cell pools and observing the ratio between the peaks of the isotopically labeled and non-isotopically labeled peptides also compensates for differential effects between the cell pools themselves, such as the presence of a different number of cells in each, providing an internal normalization between the cell samples.

To ensure that the change in the ratios is not caused by the isotopic enrichment itself, the process of Fig. 3 is preferably repeated with the other cell pool



being isotopically enriched, *i.e.*, if in the first run the treated cell pool is isotopically enriched, as in Fig. 3, then in the second run, the control cell pool would be isotopically enriched.

The ratios of the intensities of the peaks in the hypothetical spectrum shown in Fig. 4 is about 0.70, except for peptide number 3, whose ratio is about 0.30. This indicates that the relative amount of the protein corresponding to peptide 3 in the treated and control cell pools is different than the relative amounts of the protein or proteins corresponding to peptides 1, 2 and 4. This suggests that the modulation affects the post-translational expression of the protein from which peptide 3 is derived. In an actual example, there would be other pairs of peaks having ratios of about 0.30, corresponding to other peptides from the protein from which peptide number 3 is derived. Analysis of the plurality of peaks having a ratio of about 0.30 would enable an identification of the protein through the protein identification algorithms discussed above.

The percentage difference between the regularly observed ratio of the peak intensities of the peptides from both cell pools and the observed ratio for the modified peptides, *e.g.*, peptide 3, in Fig. 4, is the percentage change in the expression of the protein in the treated cell pool, *e.g.*, Cell Pool 2, which could be caused by the disease or another type of treatment or stimulus. In this example:

$$\frac{0.30 - 0.70}{0.70} \times 100 = -57\%$$

This suggests that the disease decreased the expression of the protein in the modulated Cell Pool 2 by 57%, compared with the expression of the same protein in the control Cell Pool 1. This quantitative change caused by the disease may give insight into the effect of the disease or other such modulation on the cell.

The method of the invention can also be extended to comparing the effects of two or more modulations by preparing three or more cell pools wherein at least one isotope in each cell pool is different. For example, if the proteins in three cell pools are to be compared, preferably one cell pool, which in this example is the non-modulated cell pool, would have a normal abundance of isotopes, the second cell pool would be enriched in <sup>15</sup>N and the third cell pool would be enriched in a different isotope, such as carbon-13, as well as <sup>15</sup>N. The resulting mass spectrum

- 14 -

would then comprise groups of three closely spaced peaks, each corresponding to a protein or peptide from a different cell pool. The ratios between the peak intensities of the proteins or peptides from each of the modulated cell pools and the non-modulated cell pool would be indicative of the relative amount of that peptide in each cell pool.

### Example 1

To demonstrate the internal normalization provided by mixing isotopically enriched and non-enriched proteins, the following experiment was conducted. No modulation was applied to either cell pool in this example.

Human Abelson protein tyrosine kinase Src homology domain-2 (Abl-SH2) was derived from E. coli DH $\alpha$ 5 strain cells grown in minimal medium M9, with 1 gram per liter of  $^{15}\text{NH}_4\text{Cl}$  (ammonium chloride including  $^{15}\text{N}$ ) as the sole source of  $^{15}\text{N}$ . The preparation of labeled and unlabeled Abl-SH2 from E. coli DH $\alpha$ 5 is described in Gosser, Y.Q., et al., Structure 3, 1075 (1995), which is incorporated by reference, herein. The following samples were prepared:

**TABLE II**

Sample Number	1	2	3	4	5	6	7
Natural Protein (A) picomole/microliters	0.1	0.25	0.5	1.0	1.5	2.0	2.5
$^{15}\text{N}$ Labeled Protein (B) picomole/microliters	1.5	1.5	1.5	1.5	1.5	1.5	1.5

The natural samples were prepared by dilution into 50mM ammonium carbonate ( $\text{NH}_4\text{HCO}_3$ ). 10 microliters of the natural and labeled proteins were combined and subjected to SDS-PAGE with 4-20% acrylamide concentration tris-glycine gel, available from Novex, San Diego, California, in the form of pre-cast mini-gels. A voltage of 130 volts was applied for two hours. After two hours, the gel was stained with Copper Stain from Bio-Rad Laboratories, Hercules, California.

Each protein spot or band was cut out of the gel and destained twice, for 10 minutes each. The gel pieces were then washed twice, for 15 minutes each. The gel pieces were then crushed.

0.2 micrograms of trypsin, from Boehringer Mannheim, and 50 millimolar  $\text{NH}_4\text{HCO}_3$  (ammonium carbonate) having a pH of 8 were added to the crushed gel pieces. The mixture was allowed to stand for 2 hours at 37°C. The peptides were then removed from the gel pieces by acetonitrile ( $\text{CH}_3\text{CN}$ ) at a concentration of 75%.

The solution was dried in a SpeedVac, available from Savant, Holbrook, NY. The lyophilized material was then dissolved in 7 microliters of about 50%  $\text{CH}_3\text{CN}$ , about 50%  $\text{H}_2\text{O}$  and about 0.1% trifluoroacetic acid ("TFA"). 10% of the sample solution, 0.7 microliters, was mixed with 0.7 microliters of 2-5 dihydroxybenzoic acid (gentisic acid) referred to as MALDI-MATRIX DHB, at a concentration of 60 grams per liter. MALDI-MATRIX DHB is available from Aldrich Chemical Company, Inc., Milwaukee, Wisconsin. The solution was then analyzed with a Model STR Matrix-Assisted Laser Desorption/Ionization time-of-flight mass spectrometer ("MALDI-TOF-MS"), available from PerSeptive Biosystems.

A portion of the resulting mass spectrum is shown in Fig. 5. Several pairs of peaks 110, 120, 130, 140 and 150 are shown. In each pair, the peak of the higher mass-to-charge ratio (" $m/z$ ") is indicative of a peptide of Abl-SH2 from the cell pool enriched with  $^{15}\text{N}$ , while the peak at the lower  $m/z$  ratio is indicative of a peptide of the Abl-SH2 from the non-enriched cell pool, which is predominantly  $^{14}\text{N}$ . For example, in peak pair 120, peak A is indicative of a peptide from the unlabeled cell pool while peak A' is indicative of the same peptide from the labeled cell pool. Fig. 6 is an enlarged view of the region 6 of the mass spectrum of Fig. 5, including the peaks 120 and 130 in the mass-to-charge ratio ( $m/z$ ) range of about 940-990. The corresponding peaks from Fig. 5 are similarly labeled in Fig. 6. The closely bunched multiple peaks in Fig. 6 are due to the naturally occurring distribution of the various isotopes of nitrogen, carbon, oxygen, sulfur and hydrogen in the peptides.

- 16 -

The amino acid sequence of Abl-SH2 appears below:

	1	10	20	30	40	50
1	GSGNSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQRSISLR					
51	YEGRVYHYRINTASDGKLYVSSSRFNTLAELVHHHSTVADGLITTLHYP					
101	APKRG IHRD [SEQ. ID NO: 1]					

The pair of peaks 110 m/z ~ 737 in Fig. 5 corresponds to the peptide sequence starting with the amino acid number 55 and ending with 59; the pair of peaks 120 at m/z ~ 940 corresponds to the peptide sequence spanning amino acid numbers 68-75; the pair of peaks 130 at m/z ~ 976 corresponds to the peptide sequence spanning amino acid numbers 37-45; the pair of peaks 140 at m/z ~ 1225 corresponds to the peptide sequence spanning amino acid numbers 9-18; and the pair of peaks 150 at m/z ~ 1910 corresponds to the peptide sequence spanning amino acid numbers 19-36.

The average peak intensity ratio and relative standard deviation ("RSD") for each of 5 peptides from the digested Abl-SH2 in 7 individual mass spectral measurements are shown below:

TABLE III

## Peak Intensity Ratio

Theoretical Loading Amount of  $^{14}\text{N}$  protein - 1 picomole ("pm")  
( $^{15}\text{N}$  protein amount - 1.5pm)

Peptide	m/z (M+H)	Peptide Sequence	Average Ratio (n=7)	R.S.D.
1	737	55-59	0.61	6.9%
2	940	68-75	0.60	5.6%
3	976	37-45	0.67	13.6%
4	1225	9-18	0.62	3.1%
5	1910	19-36	0.61	6.0%
5 peptides total			0.62	4.8%

The average of the RSD for the seven spectral measurements of the five peptides was 4.8%.

Fig. 7 is a graph of the measured ratios between the peak intensity corresponding to unlabeled  $^{14}\text{N}$  and the peak intensity corresponding to labeled  $^{15}\text{N}$ , based on seven spectral measurements of each of the samples of Table II, above. Since no modification was applied to either cell pool, the only difference between the samples in this example was the concentration of the natural protein in each cell pool. The value of the intensity ratios of each of the pairs of peaks for each peptide should therefore be equal to the ratio of the concentrations of the protein Abl-SH2 in each cell pool. The intensity ratio was found to be linear ( $R=0.997$ ) over an abundance ratio of more than 10:1.

### Example 2

In another example, the levels of high abundance proteins derived from two pools of *Saccharomyces* ("S.") *cerevisiae* cells that differed only in their ability to express the G1 cyclin *CLN2* was analyzed. *CLN2* is important in regulating the G1-S transition in budding yeast, but the effect of its expression on the levels of specific proteins and their modification is largely unknown. See F.R. Cross, *Curr. Opin. Cell Biol.* 6, 790 (1995). The cells in one cell pool were *cln1 cln2*, mutant yeast harboring a *GAL1::CLN2* overexpression cassette. The cells in the second cell pool were yeast without the overexpression cassette. Both populations were proliferating but only one population was expressing *CLN2*. The expressing population is designated "*CLN2*<sup>+</sup>". The non-expressing population is designated "*cln2*<sup>-</sup>".

A first combined sample contained 1 ml of unlabeled ( $^{14}\text{N}$ ) extract of *cln2*<sup>-</sup> plus 1 ml of  $^{15}\text{N}$ -labeled extract of *CLN2*<sup>+</sup>. A second combined sample contained 1 ml of unlabeled ( $^{14}\text{N}$ ) extract of *CLN2*<sup>-</sup> plus 1 ml of  $^{15}\text{N}$ -labeled extract of *cln2*<sup>-</sup>. These two different samples were prepared to control for systemic errors in the measurements.

The enriched media was Bio-Express-1000 (discussed above), specified by the manufacturer as enriched to greater than 96%  $^{15}\text{N}$ . The enriched

- 18 -

and non-enriched media were both glucose-free and were supplemented with unlabeled tryptophan. Galactose was added to each media to 0.3%. The cells were grown overnight to mid-log phase ( $OD \leq 1.0$ ) at 30° C with shaking.

The combined samples were separated by a combination of reversed phase high performance liquid chromatography ("HPLC") and SDS-PAGE. The HPLC column was 10 mm x 100 mm, and contained C4 silica gel (Brownlee Prep-10 Butyl, 20  $\mu$ m, 300 Å, Perkin Elmer, Norwalk, CT). The HPLC mobile phase A was composed of 67% ACS grade formic acid and 33% water. The mobile phase B was composed of 67% formic acid and 33% acetonitrile. The flow-rate was 1 ml/min. The gradient curve was 0 - 5 min B conc. 0%; 5 min - 10 min B conc. from 0 - 30%; 10 min - 110 min B conc. from 30 - 100%; 110 min - 120 min B conc. was 100%.

A total of 57 fractions of 2 ml each were collected. The HPLC injection volume was 14 ml, composed of 2 ml of whole yeast extract and 12 ml of mobile phase A.

After HPLC, the proteins were precipitated from each 2 ml fraction by adding 10 ml of water, 1 ml of 0.4% deoxycholate and 1 ml trichloroacetic acid (1g/ml). The solution was maintained at 4°C for 1 hr and then centrifuged at 2,200 x g for 30 min at 4°C. The supernatant was then discarded. One milliliter of 80% acetone was added and the sample was transferred to a new Eppendorf tube. This sample was maintained at -20°C for 1 hr before centrifuging at 14,000 x g for 5 min at room temperature. The supernatant was discarded and aqueous NaOH was added for neutralization.

SDS-PAGE sample buffer was added and each fraction was run on a separate lane of an 8-16% gradient tris-glycine gel at constant voltage of 110 V. The gels were stained with colloidal Coomassie Brilliant Blue. The gradient tris-glycine gel and Coomassie Brilliant Blue are available from Novex, for example.

The protein bands were excised from the gels with a scalpel and placed into Eppendorf tubes. The gel pieces were destained and washed until clear with methanol/water/acetic acid 5:4:1 (v/v/v) with 3X exchange of solution,

followed by a 30 min vortex in high purity water. The water was discarded and the gel sliced into 2-3 mm pieces, soaked in 0.5 ml of acetonitrile, and vortexed.

After the gel pieces had shrunk and turned opaque, the acetonitrile was discarded. A trypsin solution of 0.2 micrograms ("μg") in 50mM NH<sub>3</sub>HCO<sub>3</sub> was added in sufficient volume, 20-50 microliters ("μl"), to re-hydrate the gel pieces. After incubation for 2 hours at 37°C, 30 μl of acetonitrile was added to the gel pieces and vortexed for a few minutes. A further 30 μl of acetonitrile was added and the process repeated until the gel turned opaque white. The supernatant was transferred to a new tube, the gel was rehydrated with water (30 μl) and the extraction steps repeated. The supernatant was dried using a SpeedVac and each dried sample was re-dissolved in 5 μl of acetonitrile/0.1% aqueous trifluoroacetic acid ("TFA") 1:2 (v/v). 0.5 μl of this sample solution was loaded onto the sample plate together with 0.5 μl of matrix solution (DHB: 2,5-dihydroxybenzoic acid). Mass spectrum measurements were obtained using a MALDI-TOF-MS Model STR, from PerSeptive Biosystems, Framingham, MA, operated in reflector mode.

The abundances of a selection of individual proteins from the two samples were then compared. Figs. 8 and 10 are examples of MALDI-TOF-MS mass spectra of tryptic peptides from the combined cell pools of unlabeled <sup>14</sup>N *cln2*<sup>-</sup> and labeled *CLN2*<sup>+</sup>. The peptides in the mass spectrum of Fig. 8 originate from one single protein while the peptides in the mass spectrum of Fig. 10 originate from another single protein. These proteins were isolated from several hundred separated by the combination of reversed-phase HPLC and SDS-PAGE, as described above. Fig. 8 shows pairs of peptide peaks 160, 162, 164 and 166. Fig. 10 shows pairs of peptide peaks 168, 170, 172, 174 and 176. Peaks 160a, 162a, 164a and 166a in Fig. 8 and peaks 168a, 170a, 172a, 174a and 176a in Fig. 10 arise from the unlabeled (<sup>14</sup>N) *cln2*<sup>-</sup> protein. Peaks 160b, 162b, 164b, 166b in Fig. 8 and peaks 168b, 170b, 172b, 174b and 176b in Fig. 10 arise from <sup>15</sup>N labeled *CLN2*<sup>+</sup> protein.

Fig. 9a is an enlarged view of one pair of peaks 160 in region 9a of the mass spectrum of Fig. 8. The cluster of peaks 160a corresponds to isotopically resolved components of the unlabeled peptide while the cluster 160b corresponds to the isotopic components of the <sup>15</sup>N labeled peptide. Fig. 9b shows the theoretical

isotope distribution of  $^{14}\text{N}$  (160a') and  $^{15}\text{N}$  (160b'). Tests of the goodness of fit of the theoretical isotope distribution (FIG. 9b) to the experimental distribution (FIG. 9a) revealed that the level of incorporated  $^{15}\text{N}$  was  $93 \pm 1\%$ . The multiple pairs of peaks within a spectrum provide multiple measurements of the relative abundance for each identified protein. For each protein subjected to mass spectroscopy, the ratio of the abundance of the protein in the two cell pools was obtained from the intensity ratios of the pairs of peaks in the corresponding spectrum of tryptic peptides by comparing the sum of the intensities of the isotopically resolved components of the unlabeled peptide with the corresponding sum from the  $^{15}\text{N}$  labeled peptide.

The sets of masses of the lower mass components of each pair (i.e., sets of tryptic peptides from the unlabeled proteins) were used to identify the proteins from the *S. cerevisiae* database with the protein identification algorithm ProFound, discussed above. The spectrum of Fig. 8 was found to originate from the protein "elongation factor 1- $\alpha$ " ("EF1- $\alpha$ ") while the spectrum of Fig. 10 was found to originate from the protein "triosephosphate isomerase" ("TIM"). The ProFound probability value ("P") for EF1- $\alpha$  was 1.00, readily discriminating against the second ranked choice of clb5 ( $p=9.0^{-20}$ ). The ProFound probability score for TIM was 1.00, while the second ranked choice was ORF YDL100c ( $P=1.0^{-16}$ ).

The relative abundance of a selection of proteins in the *cln2*<sup>-</sup> versus CLN2<sup>+</sup> cell pools determined from the ratio of labeled ( $^{15}\text{N}$ ) versus unlabeled ( $^{14}\text{N}$ ) peptide mass spectrometric peak intensities is shown in Table IV, below.

Measurement of 42 high abundance yeast proteins revealed that these ratios fall into two categories. The first category, which includes the majority of the proteins studied, yielded intensity ratios that are the same to within the relative experimental error ( $\pm 10\%$ ) of the measurement. The average of this category of intensity ratios were normalized to 1.00 on the assumption that they arise from proteins whose relative abundances do not change in the two cell pools. Elongation factor 1- $\alpha$  falls into this first category. The second category arises from proteins whose relative abundance differs in a statistically significant manner (Standard Deviation > 3) from the first category. Triosephosphate isomerase falls into this second category because its unlabeled-to-labeled peak ratio was determined to be 0.58. Only two



- 21 -

other proteins of the 42 sampled, a putative peroxisomal membrane protein, ORF YLR109w (4) (ratio 0.67) and S-adenosylmethionine synthetase 2 (Sam2) (ratio 0.70), were observed to fall in this second category.

TABLE III

5	Molecular Mass (kDa)		Gene	Protein Name	$\frac{cln2^+ (^{14}\text{N})}{CLN2^+ (^{15}\text{N})}$	$\frac{cln2^+ (^{15}\text{N})}{CLN2^+ (^{14}\text{N})}$
	Meas	Calc				
10	23	21.6	tsa1	Thiol-specific antioxidant protein	0.89	0.80
	27	26.7	tpi1	Triosephosphate isomerase	0.58	0.59
	29	27.5	gmp1	Phosphoglycerate mutase 1	1.08	1.14
15	34	34.8	bel1	Guanine nucleotide binding protein	1.10	1.08
	37	35.6	tdh3	Glyceraldehyde 3-phosphate dehydrogenase 3	1.12	1.04
	45	44.7	pgk1	Phosphoglycerate kinase	0.98	1.09
20	49	46.7	eno2	2-phosphoglycerate dehydratase	0.98	0.94
	51	49.9	tef1	Elongation factor 1- $\alpha$	1.00	0.84
	60	54.5	cdc19	Pyruvate kinase 1	1.12	0.97
25	90	93.3	eft1	Elongation factor 2	0.91	1.00
	110	111	kgd1	$\alpha$ -ketoglutarate dehydrogenase	0.96	1.11
	120	116	yef3	Elongation factor 3	0.86	1.00

In Table III, Column 1 "Meas" is the molecular mass of the peptide as determined by SDS-PAGE, Column 2 "Calc" is the calculated molecular mass and Column 3 "Gene" is the Yeast Protein Database (4) gene name.

The normalized intensity ratios from the  $cln2^+ (^{15}\text{N})/CLN2^+ (^{14}\text{N})$  (Column 6) cell pools were found to be in agreement with those from

*cln2*<sup>-</sup> (<sup>14</sup>N)/*CLN2*<sup>+</sup> (<sup>15</sup>N) (Column 5) pools to within the statistical uncertainty of the measurement, as shown in Table III, above.

Additional examples of proteins by Yeast Protein Database gene name that were identified from the two cell pools together with their abundance ratios *cln2*<sup>-</sup> (<sup>14</sup>N)/*CLN2*<sup>+</sup> (<sup>15</sup>N) in parentheses are: *ilv5* (0.93); *grs1* (0.97); *acs2* (0.99); *por1* (1.06); *pfk2*; *cdc48*; *gdh1*; *pfk2* (0.95); *cdc48* (0.90); *gdh1* (1.06); *pet9* (1.09); *pdcl* (1.00); *YLR109w* (0.67); *fba1* (1.18); *asn2* (0.97); *cys4* (1.05); *idh1* (1.10); *rpl5* (0.99); *efb1* (0.96); *YKL056c* (1.14); *tif51a* (0.94); *rps5* (1.16); *act1* (1.12); *hvk2* (1.11); *pgi1* (1.19); *ssa1* (0.95); *fas1* (0.99); *hsc82* (0.84); *hom6* (0.96); *rpp0* (1.03); *mr2* (0.91); *bmh1* (0.91); *sam2* (0.70).

Through the process of the present invention, it was demonstrated that the abundance of the triosephosphate isomerase, ORF *YLR109w* and S-adenosylmethionine synthetase 2 (*Sam2*) was less in the *cln2*<sup>-</sup> sample than in the *CLN2*<sup>+</sup> sample, while the abundance of the elongation factor 1- $\alpha$  was the same in both samples. The biological implications of the change in the abundances of these proteins remain to be elucidated.

Based on a statistical analysis of the results, it is believed that changes in the abundance of proteins outside of two standard deviations (>20%) can be readily discerned.

The method of the present invention can also be applied to the detection of changes in the levels of protein components of incompletely separated mixtures provided that individual peptides in the MALDI-MS peptide map can be unambiguously assigned to specific proteins. Thus, for example, two 75 kDa proteins, glycyl- tRNA synthetase and acetyl coenzyme A synthetase 2, were identified by the algorithm ProFound, discussed above, which also automatically identifies proteins present as binary mixtures, and their abundance changes determined from a single gel band. These same identifications were made by subjecting a portion of the same sample from the 75 kDa band to HPLC separation followed by online electrospray ionization mass spectrometry ("ESI-MS") and ESI-tandem mass spectroscopy analysis using an ion trap mass spectrometer. D. Ashcroft et al., Electrophoresis 19, 968 (1998). The tandem mass spectroscopy

o (“MS/MS”) experiment identified proteins from the fragmentation patterns of individual peptides using the search algorithm PepFrag, which can also be accessed on the World-Wide Web at the same cite as ProFound, identified above. The corresponding MS experiment was used to obtain the intensity ratio for quantification. This combined liquid chromatography mass spectroscopy (“LC-MS”) and liquid chromatography tandem mass spectroscopy (“LC-MS/MS”) approach should also allow for the quantitative analysis of even more complex mixtures of proteins.

### Site Specific Modifications

The method of the present invention also enables the identification of the site (peptide) of a modification or modifications of a protein and quantification of the difference in the degree of the modification of the protein in the control and treated cell pools, which may give insight into the mechanism of a disease or other cellular process. A hypothetical example is discussed below.

A control cell pool and an isotopically labeled, diseased cell pool are prepared and combined and the proteins extracted and separated, as described above. A single protein of interest is removed from the gel, digested and subjected to mass spectroscopy, also as described above. Fig. 4 and Table I demonstrate exemplary results for this example, as well.

Peptides from the two cell pools that either remain unmodified or do not undergo a change in the level of modification yield pairs of peaks with a fixed ratio of intensities -- a ratio that can be used to normalize the amounts of the protein from the two cell pools. By contrast, peptides that undergo a change in their level of modification yield pairs of peaks with intensity ratios that reflect these changes.

In Fig. 4, the ratios between the intensities of each pair of peaks corresponding to the same peptide from each cell pool is substantially the same, except for peptide 3. This deviation from the norm could be indicative of a change in the peptide in the corresponding protein in the diseased cell pool, which could give insight into the mechanism of the disease. For example, such a deviation could be indicative of greater or less phosphorylation, glycosylation, acylation, etc. of that peptide in the diseased cell pool than in the normal cell pool.

- 24 -

The m/z values of the pair of peaks in the mass spectrum is indicative of the type of change in the peptide, *i.e.*, whether a phosphate, carbohydrate or other such group is bonded to the peptide or not.

Two cases are considered. In the first case, the m/z ratio indicates that no modification of the peptide has taken place, for example, the peptide has not been phosphorylated. In the second case, the m/z ratio indicates that the peptide has been modified, such as by being phosphorylated. The percent difference between the regularly observed ratio and the observed ratio for peptide 3 in the second case is the change in the percentage of those peptides in the diseased cell pool which have been modified compared with the percentage of those peptides which have been modified in the control cell pool, here:

$$\frac{0.30 - 0.70}{0.70} \times 100 = -57\%$$

In this example, fifty seven percent of the peptide 3 which was present in the control cell pool has been modified in the diseased cell pool. Hence, 43% of the peptide 3 which was present in the control cell pool has not been modified in the diseased cell pool.

Conclusions can also be drawn as to the level of the state of modification, as shown in Example 3, below.

### Example 3

In this example, the degree of phosphorylation of PAK-related Sterol 20 protein kinase ("STE 20") in normal and mutant yeast cells during a signaling cascade were compared.

Cln2-Cdc28 cyclin-dependent kinase inhibits the mating factor signal transduction pathway by interfering with the function of Ste20, which correlates with Cln2-dependent *in vivo* phosphorylation of Ste20. See, for example, Oehlen, L.J.W.M., Cross, F.R., Genes Dev. 8, 1058 (1994); Oehlen, L.J.W.M., Cross, F.R., J. Biol. Chem. 273, 25089 (1998); Wu C, Leeuw T, Leberer E, Thomas DY, Whiteway M., J. Biol. Chem. 273, 28107 (1998). The differences in phosphorylation of Ste20<sub>trunc</sub> in *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup> cell pools were monitored in

- 25 -

accordance with the method of the present invention to identify Cln2<sup>+</sup>-dependent *in vivo* phosphorylation sites in Ste20.

Plasmid pYGEX-STE20 (B3553) expresses GST-Ste20 from the *GAL1* promoter, as described in R.L. Roberts, et al., *Cell* 89, 1055-65 (1997). GST-Ste20 phosphorylation site mutants were amplified via PCR from pVTU-STE 20-based constructs using an internal oligonucleotide upstream of the BamHI site and a 3' oligo which hybridized outside of the multiple cloning site of the plasmid. The 3' oligo added an SpeI site, and PCR products were cleaved with BamHI and SpeI and transferred to B3553 cut with BamHI and XbaI to create wild type and mutant STE20 alleles.

To prepare the GST-Ste20<sub>trunc</sub> fusion proteins for mass spectroscopy, GST-Ste20<sub>trunc</sub> fusion proteins spanning residues of 496-939 of full-length Ste20 for mass spectrometer analysis were made by transforming cells (strain BOY491 (*cln2*<sup>-</sup>) or BOY493 (*CLN2*<sup>+</sup>)) with a plasmid expressing GAL1p::GST-Ste20. The cells were grown in SCGal-Ura overnight, to an optical density between 0.8 and 1.0. The cells were pelleted, washed in wash buffer of 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM EDTA, and broken with glass beads in 50 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM EDTA, 0.08% Triton-X-100, plus protease inhibitors ("TNET"). Clarified extract was incubated with glutathione agarose for 1 hr at 4°C with rotation. The agarose was pelleted, washed 3 times with TNET and the protein was eluted either with an equal volume of 2X SDS-PAGE sample buffer or with 5 mM glutathione prepared in 50 mM Tris-HCl pH 8.0. Samples eluted with glutathione were concentrated using Microcon-30 microconcentrators available from Millipore Corporation, Bedford, Massachusetts.

An SDS-PAGE gel was copper stained with Bio-Rad from Hercules, CA. The Ste20<sub>trunc</sub> band (~80 kDa) was cut out, destained, washed, digested in-gel with trypsin, and the tryptic peptides extracted. HPLC separations (Michrom UMA, Michrom BioResources, Inc., Pleasanton, CA) were made with a C8 silica gel column (Inertsil C8 (150 mm x 0.7 mm, 5µm, 300Å) GL Science, Tokyo, Japan). The eluent from the HPLC column was connected directly to the electrospray ion source. Mobile phase A was acetonitrile:water (2:98 (v/v)) containing 0.1% TFA

- 26 -

and mobile phase B was acetonitrile:water (95:5) containing 0.7% TFA. A linear gradient program was run from 0 to 60% B over a period of 60 minutes (flow-rate 20  $\mu$ l/min).

Ten percent of the sample was used for MALDI-TOF-MS analysis and the other 90% for LC-MS and LC-MS/MS analysis to identify sites which were phosphorylated. LC-MS and LC-MS/MS analysis were performed with an electrospray ion trap mass spectrometer, model LCQ, available from Finnigan MAT, San Jose, CA, operated in a mode which alternated single mass spectral ("ms") scans ( $m/z$  400 – 2000) with ms/ms scans (data dependent scan mode in which the most intense ion peak in the previous ms scan was isolated and subjected to collision-induced dissociation ("CID")). The CID energy was set at 30 and the ion injection time at 100 ms. The MALDI analysis was conducted as described above in Example 2.

MALDI-TOF-MS and LC-ESI-ion trap-MS/MS analysis of unlabeled full length Ste20 as well as a truncated form spanning residues 496-939 (Ste20<sub>trunc</sub>) identified 13 sites that were phosphorylated *in vivo*: Ser<sup>418</sup> (Ser<sup>422</sup> or Thr<sup>423</sup>), Ser 502, three sites in a tryptic peptide spanning residues 506-530, Ser<sup>547</sup>, (Ser<sup>521</sup>, Thr<sup>552</sup> or Thr<sup>555</sup>), Ser<sup>562</sup>, Thr<sup>573</sup>, Ser<sup>585</sup>, Thr<sup>773</sup>, and (Ser<sup>861</sup> or Thr<sup>863</sup>). The ambiguities in the identification of certain of the sites arise because MS/MS analysis does not always provide information on each amino acid residue in the peptide sequence.

To monitor the differences in phosphorylation of STE20<sub>trunc</sub> in *CLN2*<sup>+</sup> versus *cln2*<sup>-</sup>, two cell pools were prepared. One cell pool contained *CLN2*<sup>+</sup> in a medium containing a natural abundance of <sup>14</sup>N and the other cell pool contained *cln2*<sup>-</sup> in a medium enriched in <sup>15</sup>N. The media for each cell pool were prepared as described above in Example 2. Fractions from the two cell pools were mixed and subjected to SDS-PAGE. The band containing the mixture of labeled and unlabeled STE20<sub>trunc</sub> was excised and digested with trypsin. The resulting peptides were extracted by HPLC on a Michrom UMA instrument with a pre-column splitter and 50 mm x 0.2 mm C18 silica gel capillary column. The capillary column used was the Magicms, 200Å, 5 $\mu$ m, available from Michrom BioResources, Inc. Mobile

phase A was methanol:water (5:95) containing 1.0% acetic acid, and mobile phase B was methanol:water (85:15) containing 1.0% acetic acid. A linear gradient program was carried out from 0 to 60% of B concentration for 30 minutes. The total flow-rate was 50  $\mu$ l/min prior to pre-column splitting and  $\sim$ 3 $\mu$ l/min after splitting. The total event was directed to an electrospray ion trap mass spectrometer, which was operated in single-stage MS profile mode over the range of m/z 400 – 2000 with an ion injection time of 100 ms. The mass spectrometer was a Finnigan LCQ electrospray ion trap spectrometer, available from Finnigan MAT, San Jose, California.

Fig. 11 is a schematic representation of the resulting mass spectrum. For the purpose of illustration, peptides that remain unchanged in the two cell pools are assumed to be present in equal abundance and the level of phosphorylation of peptide X is assumed to change from 30% (pool 1) to 70% (pool 2) – leading to a decrease in the measured intensity ratio of unphosphorylated peptide X and an increase for phosphorylated peptide ( $X_p$ ). Fig. 12 shows the actual changes in the ratios for the unphosphorylated peptide X and the phosphorylated peptide  $X_p$ .

Measurement of the intensity ratios of the isotopically labeled (*cln2*<sup>+</sup>) versus unlabeled (*CLN2*<sup>+</sup>) phosphopeptides showed that at least four of these sites exhibited large increases in phosphorylation in the *CLN2*<sup>+</sup> cell pool. These *Cln2*-dependent sites appear to be consensus cyclin dependent S/T-P sites, consistent with direct phosphorylation of Ste20 by *Cln2*-Cdc28.

It was found that the Ste20 peptide SKTSPIISTAHTPQQAQSPK<sup>564</sup> [Sequence I.D. No. 3] was phosphorylated at Ser<sup>547</sup>, Ser<sup>562</sup> and (Thr<sup>551</sup>, Ser<sup>552</sup>, or Thr<sup>555</sup>). The first two sites are within SP motifs (underlined) while the third site was constrained to a 5-residue stretch that contains a TP motif (underlined). Fig. 13, Box A shows the actual change in the intensity ratio of the unphosphorylated peptide X as well as the change observed for the singly, doubly and triply phosphorylated Ste20 peptide obtained from *CLN2*<sup>+</sup> versus *cln2*<sup>+</sup> cells. The intensity ratio for the unphosphorylated peptide X decreased by  $84 \pm 5$  % while that for the singly, doubly, and triply phosphorylated sites increased by  $24 \pm 12$  %,  $44 \pm 14$  %, and  $>150$  %, respectively, showing that phosphorylation is enhanced at all three sites in

the CLN2<sup>+</sup> versus the *cln2<sup>-</sup>* mutant cells. By contrast, the Ste20 phosphopeptide SLSKELNEK<sup>591</sup> [Sequence ID. No. 4] (phosphorylated on Ser<sup>585</sup>) undergoes no significant ratio change ( $9 \pm 13$  %, Fig. 13, Box B), demonstrating that phosphorylation at Ser<sup>585</sup> is not Cln2-dependent -in accord with the absence of a proline-directed kinase phosphorylation motif. Cln2-dependent phosphorylation was also observed in the peptide <sup>565</sup>APAQETVTTPTSKPAQAR<sup>582</sup> [Sequence ID. No. 5] (Fig. 13, Box C) and to a lesser extent in the peptide <sup>772</sup>TTMVGTPYWMAPEVVSR<sup>788</sup> [Sequence ID. No. 6] (Fig. 13, Box D). Using MS/MS analysis, we found the former to be phosphorylated on Thr<sup>573</sup> (a TP motif), while the latter peptide was phosphorylated on Thr<sup>773</sup> (and not Thr<sup>777</sup>Pro<sup>778</sup>, as was previously reported in Wu et al., J. Biol. Chem. 270,15984 (1995)). These data demonstrate that the present method can precisely discern site-specific changes in the degree of phosphorylation of a protein.

Any post-translational modification of a protein which will effect the mass of the protein may be similarly compared.

The method of the present invention is applicable to a wide range of areas. The effective quantification of protein expression levels using the method of the present invention aids in understanding the interaction of gene expression with external factors in producing phenotypes. In combination with, and as an extension of genomic transcriptional expression mapping, the quantitative description of the protein phenotype assists in understanding the molecular basis of physiological and pathological processes.

In accordance with the present invention, changes in the post-translational expression of a protein or proteins with time may be studied by periodically withdrawing samples from a control cell pool and a cell pool whose metabolism has been arrested, either one of which may be isotopically labeled.

The effects of chemical compounds on the post-translational expression of proteins in one cell pool may be compared with a cell pool which has not been exposed to the compound. Such a procedure could be useful in screening drug candidates by giving an indication of side effects. For example, one cell pool can be treated with a compound which is a potential drug candidate. Comparison



with an untreated control cell pool in accordance with the present invention can indicate the changes in post-translational protein expression caused by the drug. If it is found that the drug causes a known change in the expression of protein or proteins which is known to be deleterious, that compound can be eliminated from further development prior to the commitment of significant resources. Similarly, if it is known that a change in the expression of a protein or proteins has a positive effect, a comparison of those protein levels in the treated and control pool in accordance with the method of the invention can give a preliminary determination of whether a particular compound may be a useful drug.

With a knowledge of the deleterious and positive changes in protein expression, the toxicology of pesticides, chemicals and environmental agents can also be examined in accordance with the method of the present invention. In the field of agriculture, the effects of fertilizers, pesticides and pheromones on the post-translational synthetic expression of proteins can be studied.

The effects of gene therapy can also be studied by the methods of the present invention. One or more genes in the cells in one cell pool can be inserted, replaced, modified, overexpressed or underexpressed, as is known in the art. In accordance with the present invention, a control cell pool in which the genes have not been subjected to gene therapy is provided, as well. Either one of the cell pools can be cultured in an isotopically enriched or depleted medium while the other is cultured in a medium having naturally occurring isotopic proportions. The method of the present invention enables the effects of such a genetic modification on post translational protein expression or other cellular functions to be studied.

The effects of treatments on the cell membranes themselves can also be studied in accordance with the present invention. In this case, isotopically enriched and non-isotopically enriched cell pools are combined, the cell membranes are removed from the combined cell pool in a manner known in the art, digested and subjected to mass spectroscopy.

The secreted by-products of the cell pools could also be compared by drawing samples from the media of each pool, one of which is isotopically enriched, mixing the samples, removing the cells and any other unwanted components, and

- 30 -

analyzing the remaining mixture of the sample media by mass spectroscopy, as described above.

The post-translational effects of hormones, infectious agents such as viruses and bacteria, carcinogens, and trauma, such as burns, can be similarly studied and quantified. Pain modulation can also be examined.

The effects of cell differentiation on post-translational protein expression can also be studied in accordance with the method of the present invention.

The applications described herein are merely examples of several of the many possible uses of the process of the present invention. While the present invention has been described with respect to quantifying post-translational changes in protein expression, changes in protein expression resulting from the effects of any modulation on translation or transcription can also be studied and quantified. In addition, the process of the present invention may be used to compare the relative quantities of any biological component which can be ionized or whose subcomponents can be ionized so that it can be analyzed by mass spectroscopy, in any type of biological matter which can be grown in an isotopically labeled medium. For example, in addition to the proteins and peptides discussed above, the biological component may be a nucleic acid, a carbohydrate, a lipid, a cofactor and post-synthetic derivatives thereof.

In addition to the biological cells discussed above, the biological matter itself may be a microbiological culture, biological tissue, an organ, an organism, a collection of organisms, a part of an organism, and a cell-free biological mimetic system, for example.

The biological component can also act as a marker for effects on biological processes. For example, if it is known that the level of a protein changes due to a change in a biological process caused by a modulation, the change in that protein level can be used to study the response of the biological process to the modulation even if the role of that protein in that process is not understood. A marker could be any biological component.

As an example of the comparison of the effects of a modulation on the characteristics of organisms, two cultures of nematodes can be grown, one in a medium containing a normal abundance of isotopes and the other in a medium isotopically enriched or depleted in at least one isotope. One of the cultures may then be modulated. The cultures can then be mixed, the nematodes broken up, such as by crushing, and a portion of the combined culture extracted and subjected to mass spectroscopy to analyze the differential effect of the modulation on the nematodes from each cell pool.

Cells, tissue, fluids or other biological matter may also be withdrawn from a human or animal subject fed isotopically enriched or depleted food for comparison with the same biological matter withdrawn from another human or animal subject fed food having normal isotopic proportions, and analyzed in accordance with the present invention. Biological matter withdrawn from different species of animals can also be compared in accordance with the present invention to study the differences in protein expression, and other differences, between the species.

Samples from the same human or animal subject can be analyzed at different points in time, as well. First, biological matter such as cells, tissues or fluids are withdrawn from the subject. Then the subject is fed food enriched in one or more isotopes. After sufficient time for metabolism, the same biological matter is withdrawn from the subject. The withdrawn materials are mixed and analyzed as described above. Samples can be withdrawn at multiple times to monitor the metabolism of the food with time, as well. Such analyses may be useful in clinical investigation and diagnosis.

As mentioned above, one of the samples of biological matter need not be cultured, grown or maintained in a medium or food having a normal abundance of isotopes. As long as the abundance of at least one isotope in one of the media or the food fed one of the animals is different from the abundance of that isotope in the other, the method of the present invention may be applied.

We claim:

o

1. A method for comparing the relative abundance of a protein of interest in multiple samples of biological matter comprising:

culturing a first sample of biological matter in a first medium containing a natural abundance of isotopes;

5

culturing a second sample of biological matter in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium;

modulating one of the samples;

10

combining at least a portion of each of the samples;

removing at least one protein from the combined samples;

subjecting the at least one removed protein to mass spectroscopy to develop a mass spectrum;

15

computing a ratio between the peak intensities of at least one pair of closely spaced peaks;

determining the relative abundance of the protein in each sample based on the at least one computed ratio; and

20

identifying the protein.

2. The method of claim 1, further comprising identifying the protein by electrophoresis, antibodies, bioassay or from the mass spectrum.

25

3. The method of claim 1, wherein the removing step comprises extracting proteins from the combined cell pool, the method further comprising separating the extracted proteins by a process chosen from the group consisting of one-dimensional electrophoresis, two-dimensional electrophoresis, ultra-

30

centrifugation, chromatography and affinity binding.

4. The method of claim 3, wherein the separating step comprises:

35

subjecting the extracted proteins to electrophoresis in a gel; staining the gel;

- 33 -

removing gel spots of interest; and  
destaining the gel spots of interest;  
the method further comprising digesting the proteins in the  
destained gel spots.

5                    5.     The method of claim 1, further comprising digesting the at  
least one protein of interest.

10                   6.     The method of claim 1, further comprising computing ratios  
of the peak intensities of a plurality of pairs of closely spaced peaks of the mass  
spectrum.

15                   7.     The method of claim 1, further comprising:  
removing a plurality of proteins of interest from the combined  
cell pool;  
digesting the plurality of proteins of interest into a plurality of  
peptides;  
subjecting the digested proteins to mass spectroscopy to  
20     develop the mass spectrum;  
selecting a plurality of pairs of closely spaced peaks on the  
mass spectrum;  
computing the ratio of the intensities of the peaks in each  
25     pair;  
determining the protein from which the pair of peaks in the  
mass spectrum are derived based on the mass spectrum; and  
determining the relative abundance of the protein in each cell  
30     pool.

8.     The method of claim 1, further comprising determining the  
relative quantity of a modified protein of interest in each cell pool.

35

9. The method of claim 8, further comprising determining the site of the modification on the protein.

10. The method of claim 1, wherein the modulating step comprises subjecting the biological matter in the one sample to an environmental or chemical stimulus.

11. The method of claim 1, wherein the modulating step comprises genetically manipulating the biological matter in the one sample.

12. The method of claim 11, wherein the genetic manipulation comprises performing gene therapy.

13. The method of claim 1, wherein the modulating step comprises subjecting one of the cell pools to a drug or hormone.

14. The method of claim 1, wherein the modulating step comprises subjecting the cell pool to a treatment chosen from the group consisting of a virus, a bacteria and a carcinogen.

15. The method of claim 1, wherein the protein is a marker for the effect of the modulating step on a biological process.

16. The method of claim 1, comprising culturing the second sample in a second medium in which the at least one isotope is non-radioactive.

17. The method of claim 1, comprising culturing the second sample in a second medium enriched in the at least one isotope.

18. The method of claim 1, comprising culturing the second sample in a second medium depleted in the at least one isotope.

19. The method of claim 1, wherein the biological matter is chosen from the group consisting of biological cells, biological tissue, an organ, an organism, a collection of organisms, a portion of an organism, and a cell-free biological mimetic system.

20. The method of claim 1, comprising culturing the second sample in a second medium wherein at least one isotope chosen from the group consisting of nitrogen-15, carbon-13, oxygen-17, oxygen-18, sulfur 34 and hydrogen-2, has a different abundance than the abundance of the same isotope in the first medium.

21. The method of claim 20, comprising culturing the second sample in a second medium enriched to at least about 90% in nitrogen-15.

22. A method for comparing the relative abundance of a peptide of interest derived from the same protein from multiple samples of biological matter comprising:

culturing a first cell pool in a first medium containing a natural abundance of isotopes;

culturing a second cell pool in a second medium isotopically enriched in at least one non-radioactive isotope;

modulating one of the cell pools;

combining at least a portion of each of the cell pools;

extracting the proteins from the combined cell pools;

separating the extracted proteins into a plurality of proteins;

digesting at least one protein of interest from the separated proteins into a plurality of peptides;

subjecting the digested peptides to mass spectroscopy to develop a mass spectrum; and

determining the relative abundance of the peptide of interest from each sample based on the mass spectrum.

- 36 -

23. The method of claim 22, further comprising identifying the protein from which the peptide is derived and determining the relative quantity of the protein in each cell pool.

24. The method of claim 22, further comprising determining the relative occurrence of a modification of the peptide of interest in each cell pool.

25. The method of claim 24, wherein the modification of the peptide is chosen from the group consisting of the phosphorylation, glycosylation, and acylation of the peptide, the determining step comprising determining the difference in the relative abundance of the modified peptide in each cell pool.

26. The method of claim 24, further comprising determining the level of modification of the peptide.

27. The method of claim 24, wherein the type of peptide modification is chosen from the group consisting of the phosphorylation, glycosylation, and acylation of the peptide, the determining step comprising determining the difference in the relative abundance of the modified peptide in each cell pool.

28. The method of claim 22, wherein the separating step comprises subjecting the extracted proteins to a process chosen from the group consisting of one-dimensional electrophoresis, two-dimensional electrophoresis, ultra-centrifugation, chromatography and affinity bonding.

29. The method of claim 28, wherein the separating step comprises:

subjecting the extracted proteins to electrophoresis in a gel;  
staining the gel;

removing gel spots of interest; and

destaining the gel spots of interest; and



- 37 -

the digesting step comprises digesting the destained gel spots  
of interest.

30. The method of claim 29, wherein the digesting step comprises  
mixing the destained gel spots of interest with a proteolytic enzyme.

5

31. The method of claim 30, wherein the proteolytic enzyme is  
trypsin.

32. The method of claim 29, comprising culturing the second cell  
pool in a medium enriched in an isotope chosen from the group consisting of  
nitrogen-15, carbon-13, oxygen-17, oxygen-18, sulfur-34 and hydrogen-2.

10

33. A method for comparing the relative abundance of an  
ionizable component of biological matter from multiple samples, comprising:  
culturing a first sample of the biological matter in a first  
medium;

15

culturing a second sample of the biological matter in a second  
medium, wherein at least one isotope in the second medium has a different  
abundance than the abundance of the same isotope in the first medium;

20

modulating the biological matter in one of the samples;  
combining at least a portion of each of the samples;  
subjecting at least a portion of the combined sample to mass  
spectroscopy to develop a mass spectrum; and

25

determining the relative abundance of at least one component  
of interest in each sample based on the mass spectrum.

30

34. The method of claim 33, wherein the at least one component  
of interest is chosen from the group consisting of a protein, a peptide, a nucleic acid,  
a carbohydrate, a lipid, a co-factor and post-synthetic derivatives thereof.

35

35. The method of claim 33, wherein the biological matter is chosen from the group consisting of biological cells, biological tissue, an organ, an organism, a collection of organisms, a portion of an organism, and a cell-free biological mimetic system.

36. The method of claim 33, wherein the determining step comprises computing ratios of the peak intensities of at least one pair of closely spaced peaks corresponding to the at least one component of interest of the mass spectrum.

37. The method of claim 33, further comprising culturing a third sample of the biological matter in a third medium, wherein at least one isotope has a different abundance than the abundance of the same isotope in the first and second media, and modulating the third sample by a different modulation than the modulation in the first modulating step.

38. The method of claim 33, wherein the determining step comprises computing ratios of the peak intensities of at least one pair of closely spaced peaks corresponding to the at least one component of interest of the mass spectrum.

39. The method of claim 33, comprising culturing the first sample in a first medium containing a natural abundance of isotopes.

40. A method for comparing the relative abundance of a cellular component of interest in multiple cell pools, wherein the cellular component is ionizable, the method comprising:

culturing a first cell pool in a first medium;  
culturing a second cell pool in a second medium, wherein at least one isotope in the second medium has a different abundance than the abundance of the same isotope in the first medium;  
modulating one of the cell pools;

- 39 -

- combining at least a portion of each of the cell pools;  
removing at least one component of interest from the  
combined cell pool;  
subjecting the at least one removed component of interest to  
mass spectroscopy to develop a mass spectrum;  
5 computing ratios of the peak intensities of at least one pair of  
closely spaced peaks corresponding to the at least one component of interest of the  
mass spectrum; and  
identifying the component of interest.

10

41. The method of claim 40, wherein the cellular component of interest is chosen from the group consisting of a protein, a peptide, a nucleic acid, a carbohydrate, a lipid, a cofactor, and post-synthetic derivatives thereof.

15

42. The method of claim 41, wherein the component of interest is an excreted by-product of the biological matter, the method further comprising removing the biological matter from the combined cell pool and subjecting at least a portion of the medium of the combined sample to mass spectroscopy.

20

43. A method for comparing the relative abundance of biological matter from two animal subjects, wherein the biological matter can be analyzed by mass spectroscopy, comprising:

25

feeding a first animal food;

feeding a second animal food including at least one isotope having a different abundance than the same isotope in the food fed the first animal;

withdrawing a sample of biological matter from the first

30

animal;

withdrawing a sample of biological matter from the second animal;

combining at least a portion of the withdrawn biological matter from the first and second animals;

35

- 40 -

- o                                   subjecting the combined biological matter to mass spectroscopy to develop a mass spectrum; and
- determining the relative abundance of at least one component of interest of the combined biological matter based on the mass spectrum.

5                           44.     The method of claim 43, comprising feeding a second animal from a species different than the species of the first animal.

                                  45.     The method of claim 43, comprising providing first and second human subjects.

10

                                  46.     A method for comparing the relative abundance of biological matter from a single animal subject at different times, wherein the biological matter can be analyzed by mass spectroscopy, comprising:

15

                                  withdrawing a first sample of biological matter from the subject;

                                  feeding the subject food including at least one isotope having a different abundance than the same isotope in the food fed the first animal, after withdrawing the first sample;

20

                                  withdrawing a sample of biological matter from the subject after feeding;

                                  mixing at least portions of the first and second samples;

25                           subjecting the combined samples to mass spectroscopy to develop a mass spectrum; and

                                  determining the relative abundance of at least one component of interest of the combined biological matter based on the mass spectrum.

30                           47.     The method of claim 46, comprising providing a human subject.

35

- 41 -

- o 48. The method of claim 46, further comprising withdrawing at least one subsequent sample of biological matter from the subject a period of time after withdrawing the second sample, mixing at least a portion of the at least one subsequent sample with a portion of the first sample and then performing the subjecting and determining steps to monitor the metabolism of the food with time.

5

10

15

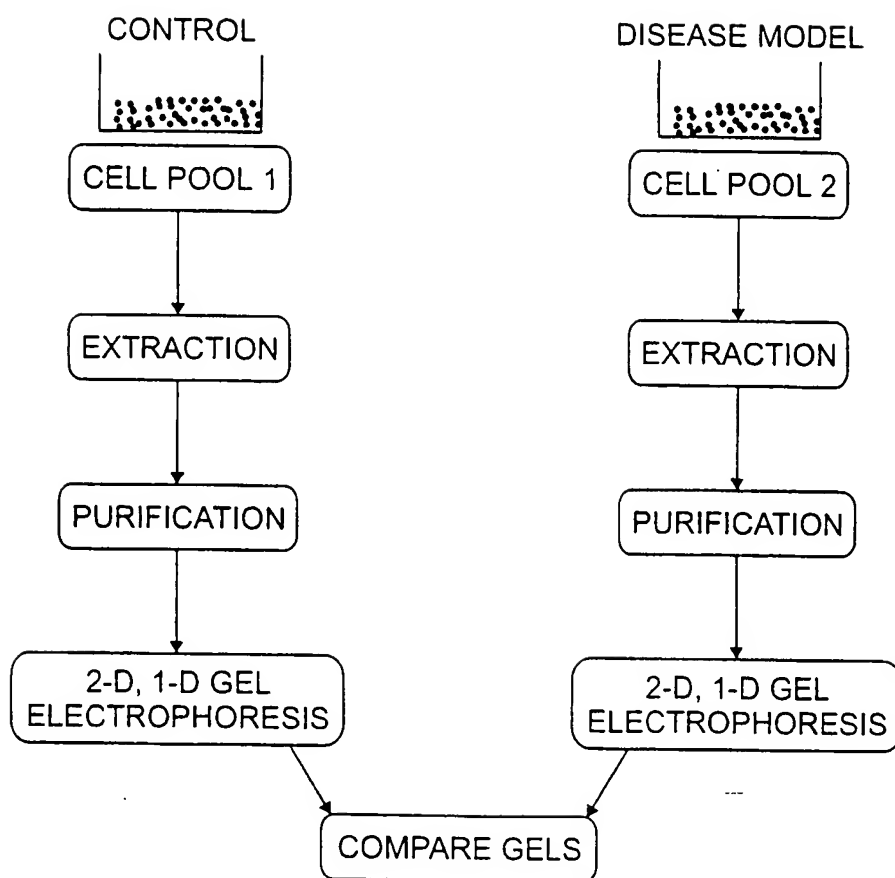
20

25

30

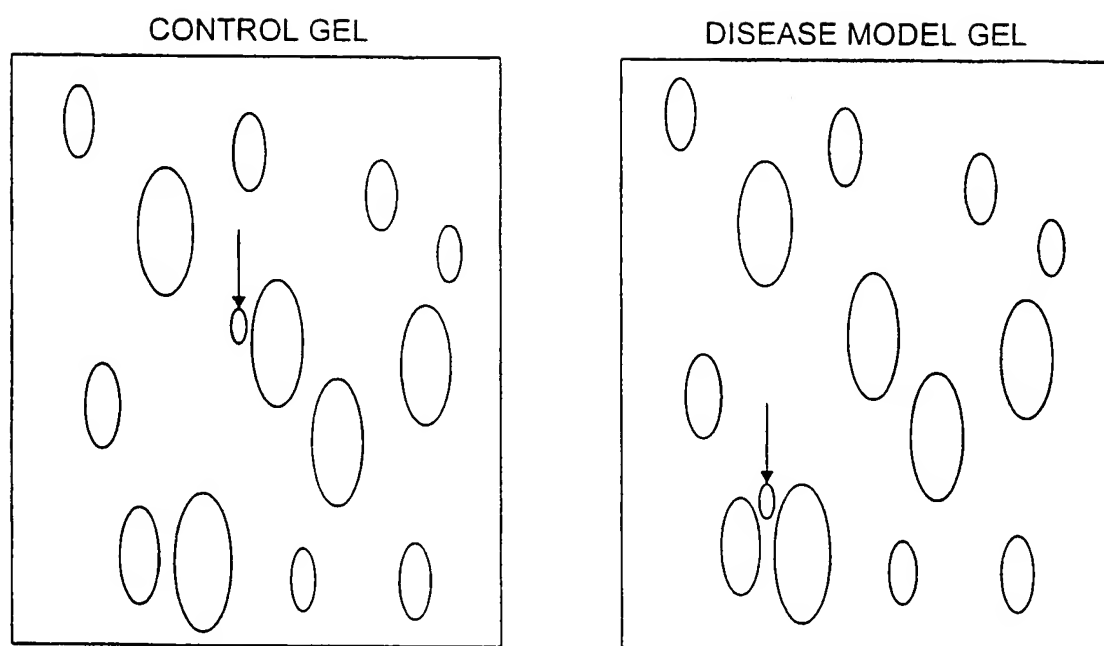
35

1/11



**FIG. 1**  
(PRIOR ART)

2/11



**FIG. 2**  
(PRIOR ART)

3/11

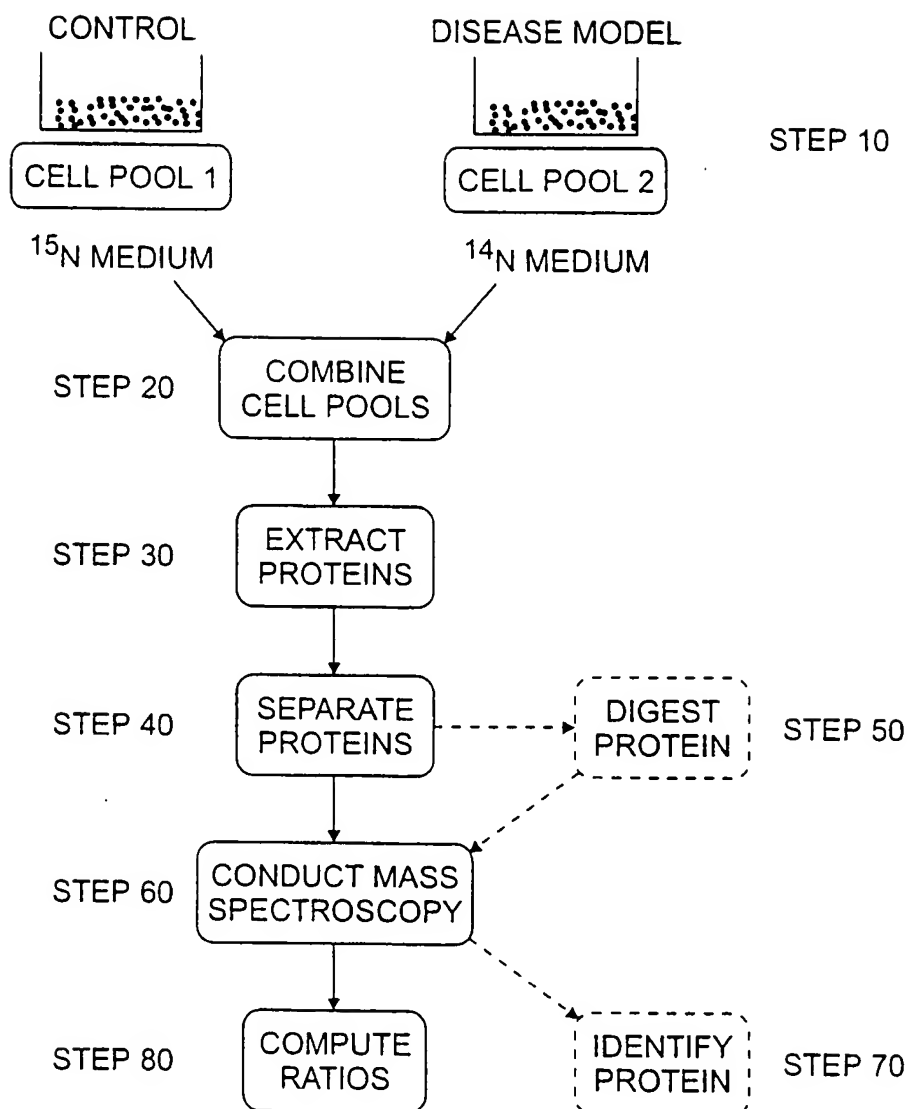


FIG. 3



4/11

$X_1^1$  = PEPTIDE i FROM  
CELL POOL 1

$X_1^2$  = PEPTIDE i FROM  
CELL POOL 2

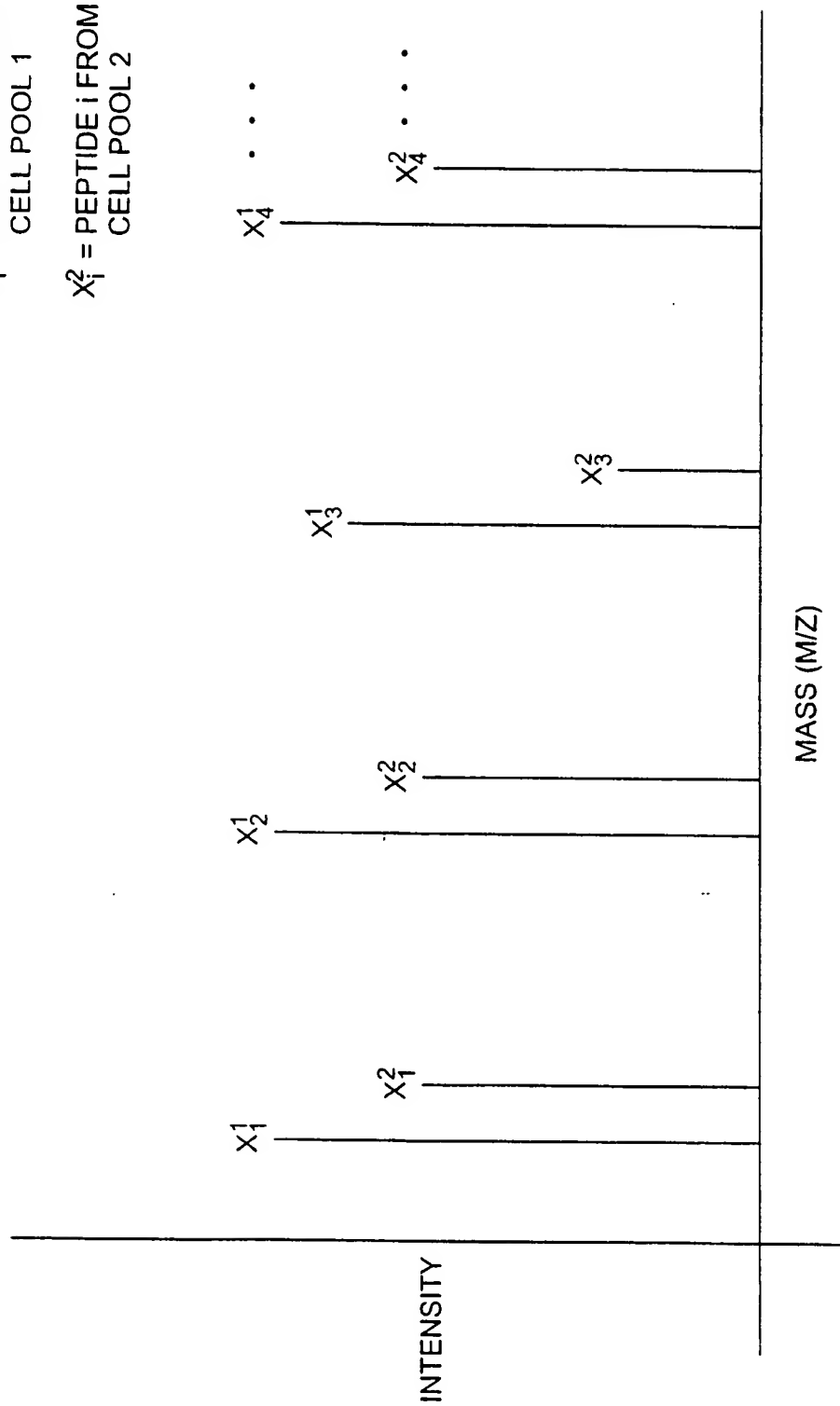


FIG. 4

5/11

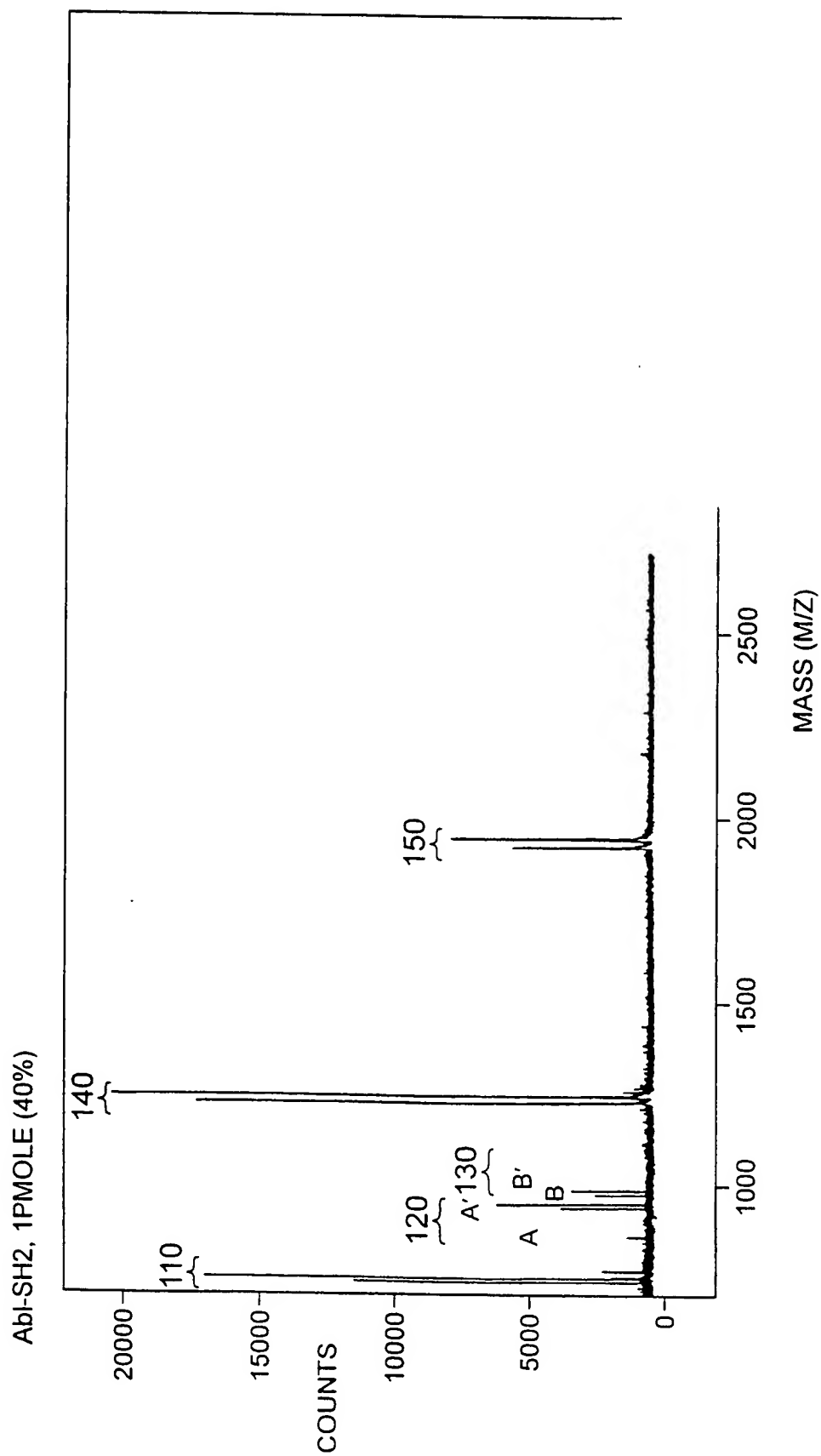


FIG. 5

6/11

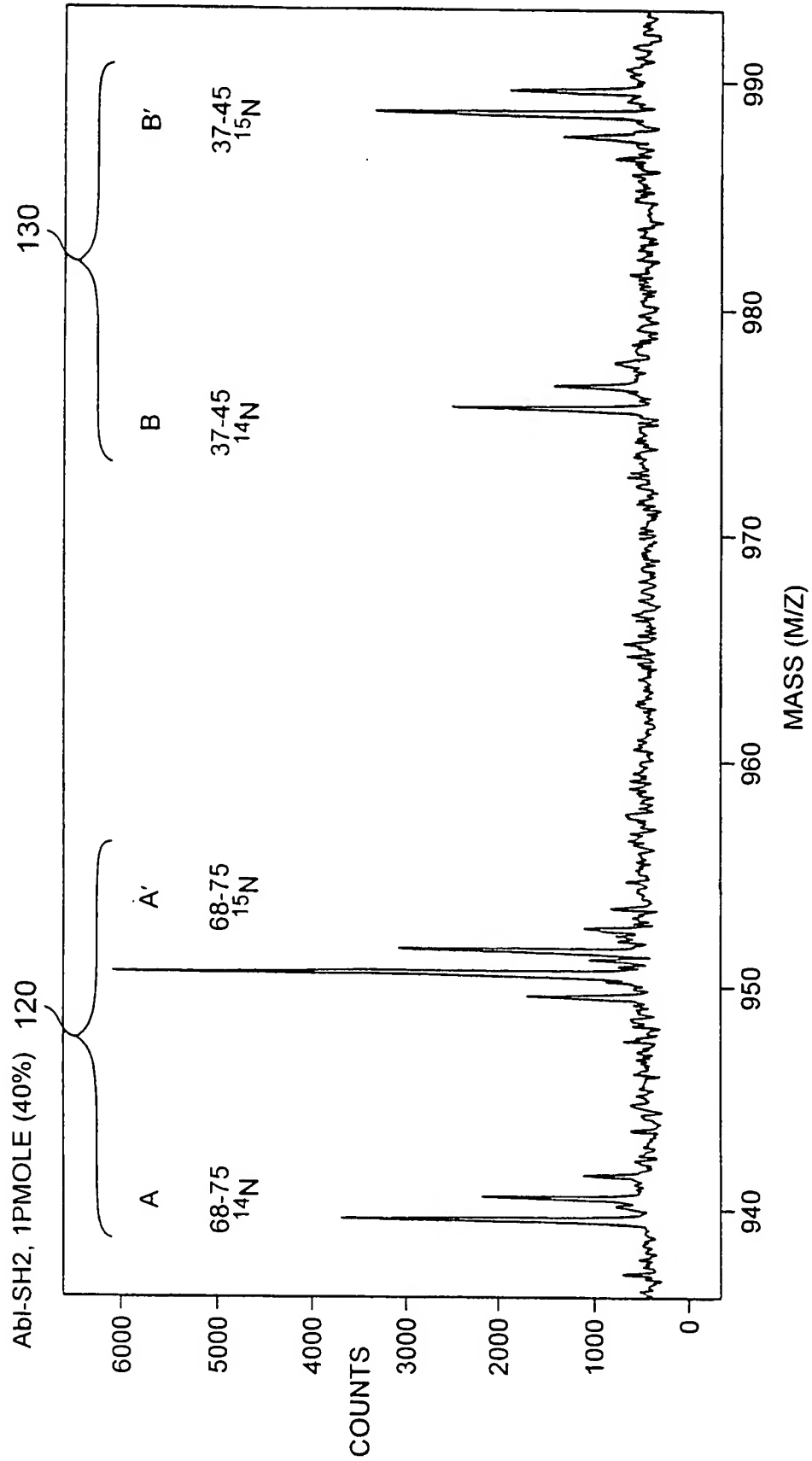


FIG. 6

7/11

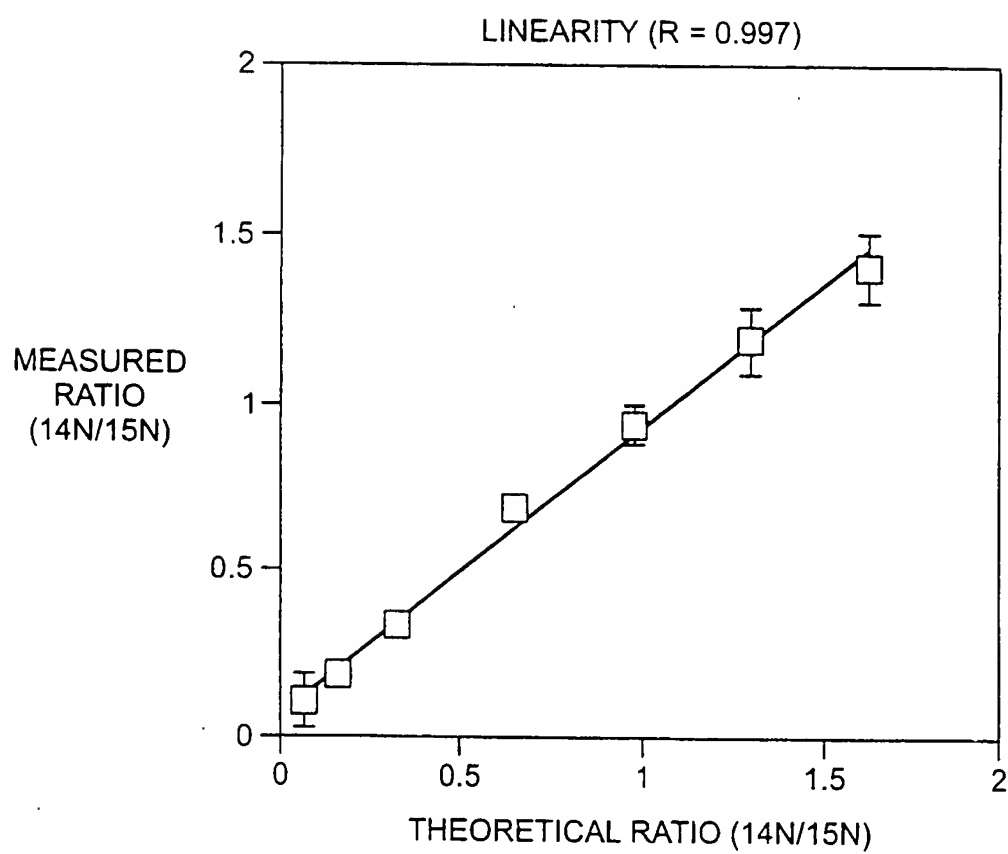


FIG. 7

8/11

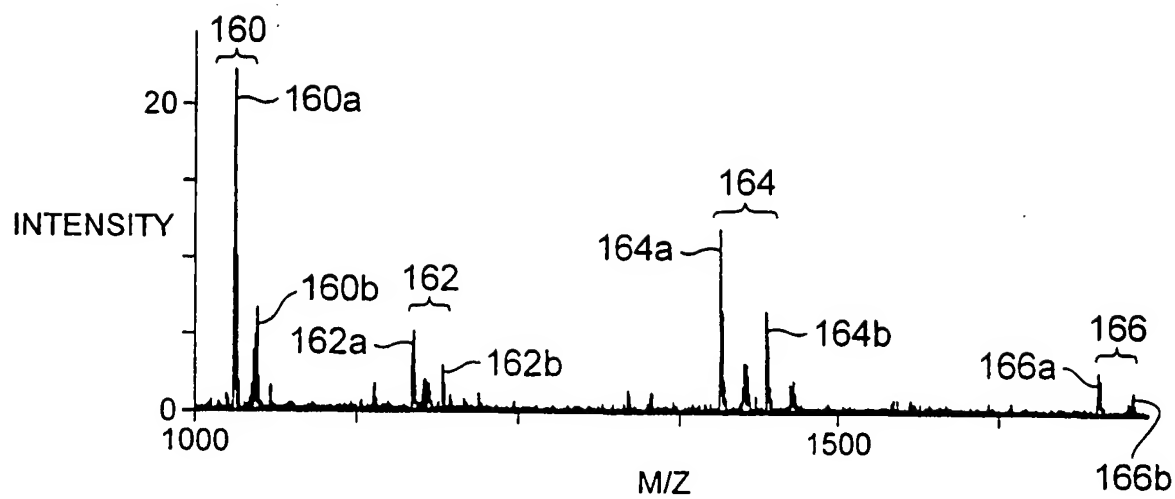


FIG. 8

FIG. 9a

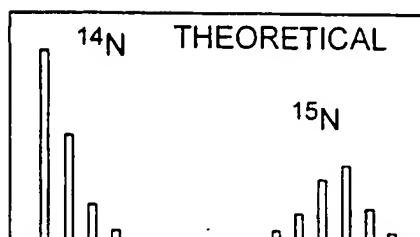
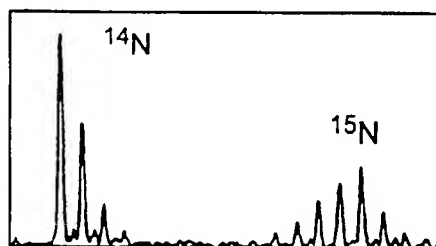


FIG. 9b

9/11

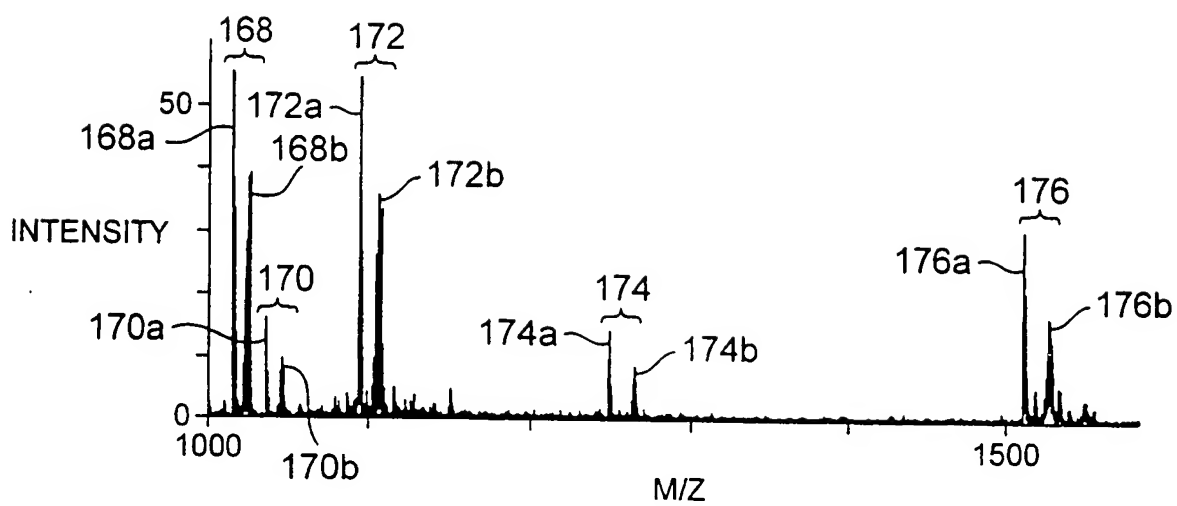
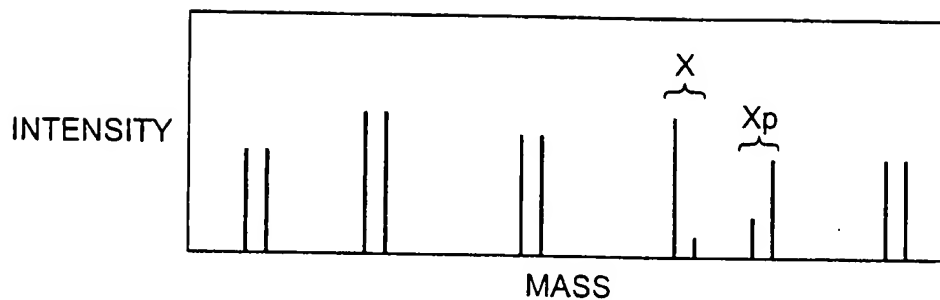
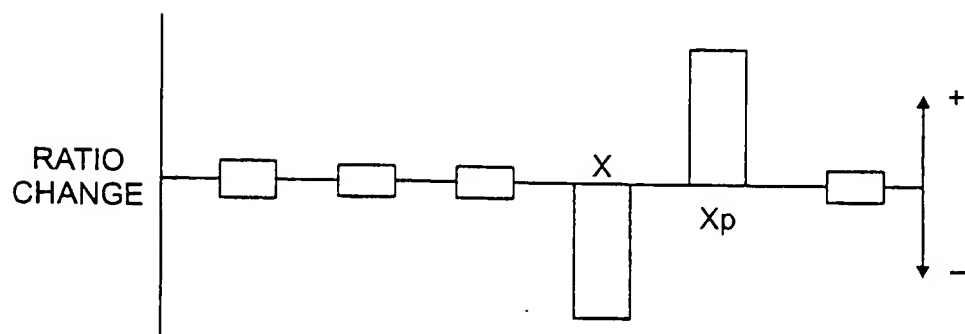


FIG. 10

10/11

**FIG. 11**

X: UNPHOSPHORYLATED PEPTIDE X  
Xp: PHOSPHORYLATED PEPTIDE X

**FIG. 12**

11/11

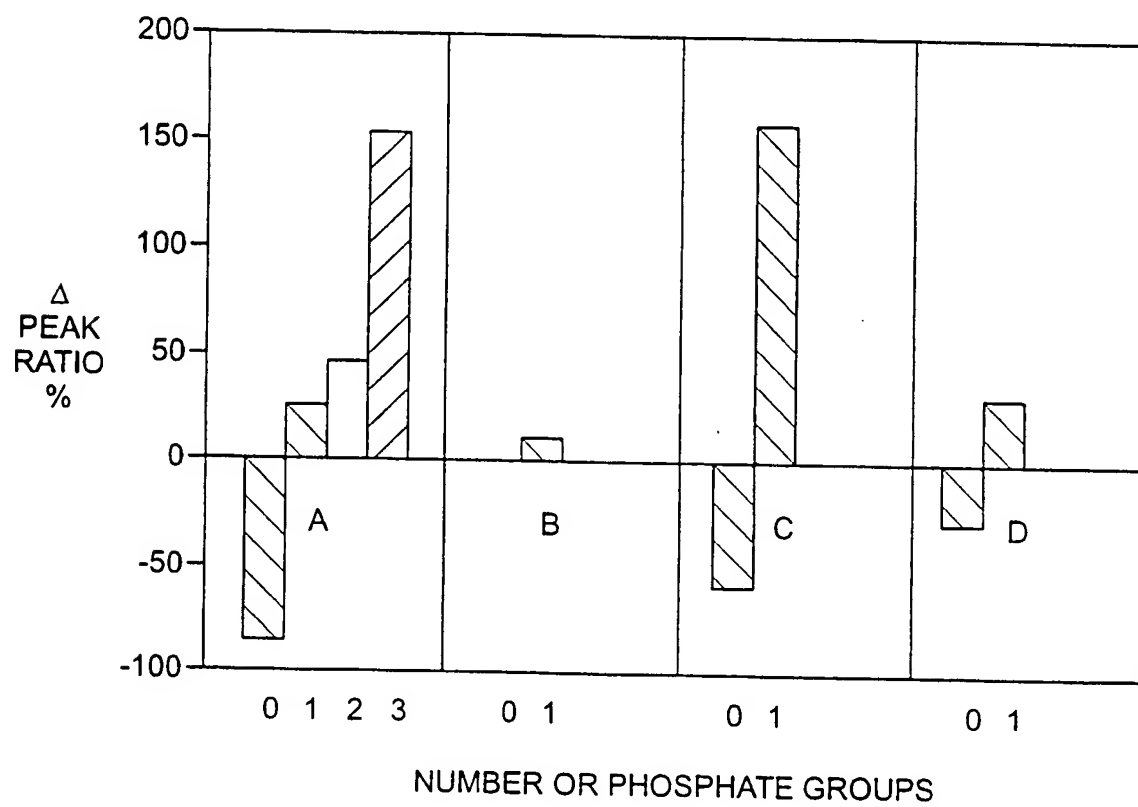


FIG. 13



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US00/12026

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G01N 33/00, 24/00

US CL : 435/71.1; 436/86,173

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/71.1; 436/86,173

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,800,979 A (KOLHOUSE et al) 01 September 1998, see entire document.	1-48
A	US 5,572,024 A (GRAY et al) 05 November 1996, see entire document.	1-48

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

14 AUGUST 2000

Date of mailing of the international search report

28 AUG 2000

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

L. BLAINE LANKFORD

Telephone No. (703) 308-0196